

# A Coding Theorem for Bipartite Unitaries in Distributed Quantum Computation

Eyuri Wakakuwa<sup>§†</sup>, Akihito Soeda<sup>\*</sup>, Mio Murao<sup>\*†</sup>

<sup>§</sup>Graduate School of Information Systems, The University of Electro-Communications, Japan

<sup>\*</sup>Department of Physics, Graduate School of Science, The University of Tokyo, Japan

<sup>†</sup>Institute for Nano Quantum Information Electronics, The University of Tokyo, Japan

<sup>†</sup>Email: wakakuwa@quest.is.uec.ac.jp

**Abstract**—We analyze implementations of bipartite unitaries by means of local operations and classical communication (LOCC) assisted by shared entanglement. We employ concepts and techniques developed in quantum Shannon theory to study an asymptotic scenario, in which two distant parties perform the same bipartite unitary on infinitely many pairs of inputs, generated by an i.i.d. (independent and identically distributed) quantum information source. We analyze the minimum cost of entanglement and classical communication per copy. For two-round LOCC protocols, we derive a single-letter formula for the minimum cost of entanglement and classical communication, under an additional requirement that the error converges to zero faster than  $1/n^4$ , where  $n$  is the number of input pairs. The formula is given by the “Markovianizing cost” of a tripartite state associated with the unitary, which can be computed by a finite-step algorithm. We also derive a lower bound on the minimum cost of resources, which applies for protocols with arbitrary number of rounds.

## I. INTRODUCTION

Distributed quantum computation is a task in which a group of distant parties collaborates to perform a large quantum computation, by using classical communication, quantum communication and shared entanglement as resources. One of the most extensively investigated tasks is implementation of bipartite unitaries by local operations and classical communication (LOCC) assisted by shared entanglement. Here, two distant parties, say Alice and Bob, have quantum systems  $A$  and  $B$  in an unknown state  $|\varphi\rangle^{AB}$ , and aim to perform a known unitary  $U^{AB}$  by LOCC using some resource entanglement shared in advance. Although this task can be implemented simply by using quantum teleportation, it was shown that the cost of entanglement and classical communication can be reduced by constructing a more efficient protocol, depending on the unitary to be implemented [1].

The following two questions then naturally arise: (i) How can we find efficient protocols which consume less resources for a given bipartite unitary? and (ii) What are the minimum cost of resources required for implementing that unitary? Although these questions have been addressed, e.g. in [1]–[11], most of the studies so far assume particular forms of the resource entanglement or of the bipartite unitary to be implemented. A general method to address these problems is yet discovered.

In the present paper, we address the above questions in an information theoretical scenario for the first time, by apply the concept of “block coding”. Here, the two parties perform the same bipartite unitary at once, on all of a sequence of input pairs generated by a completely random i.i.d. (independent and identically distributed) quantum information source. We consider an asymptotic limit of infinite pairs and vanishingly small error, and analyze the minimum cost of entanglement and classical communication per copy required for the task. We mainly focus on protocols consisting of two-round LOCC as the first nontrivial case. Our approach is different from previous approaches which have only dealt with single-shot cases [1]–[11].

The main result of this paper is that we derive a single-letter formula for the minimum cost of entanglement, forward and backward classical communication in two-round protocols, under an additional assumption that the error converges to zero faster than  $1/n^4$ , where  $n$  is the number of input pairs. The formula is represented in terms of the “Markovianizing cost” ([12]–[14]) of a state associated with the unitary, which can be computed by a finite-step algorithm. The result is applicable for any bipartite unitary.

It is left open, however, whether the same converse bound holds when we drop the requirement on the convergence speed of the error. We relate this problem to another open problem regarding an “asymptotic symmetry” of approximate recoverability, that is, whether a tripartite quantum state  $\rho^{ABC}$  is approximately recoverable from  $\rho^{BC}$  if it is approximately recoverable from  $\rho^{AB}$ , up to a dimension-independent rescaling of error of recovery. We prove that an affirmative answer to the latter question implies an affirmative one to the former.

We also derive a lower bound on the minimum cost of entanglement and classical communication, which is applicable for any protocol with arbitrary number of rounds, in terms of a parameter called the *Schmidt strength* of the unitary. It turns out that the lower bound is achievable for a class of bipartite unitaries called generalized Clifford operators.

The structure of this paper is as follows. In Section II, we introduce the formal definition of the problem. The results are summarized in Section III. In Section IV, we review results on Markovianization and state merging. Section V analyzes single-shot two-round protocols for implementing a bipartite unitary. Outlines of the proofs of the main result are presented

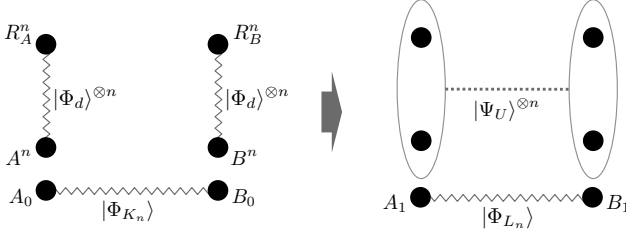


Fig. 1. The task is to apply  $(U^{AB})^{\otimes n}$  on  $(|\Phi_d\rangle^{A_0 A_1} |\Phi_d\rangle^{B_0 B_1})^{\otimes n}$  by using resource entanglement  $|\Phi_{K_n}^{A_0 B_0}\rangle$ .  $R_A$  and  $R_B$  are reference systems that Alice and Bob cannot access. The entanglement cost is defined as the difference between the amount of initial entanglement and that of final entanglement shared by Alice and Bob, i.e.,  $K_n$  and  $L_n$ .

in Section VI. In Section VII, we discuss general properties of the Markovianizing cost of unitaries as examples. The Markovianizing cost for two classes of bipartite unitaries is computed in Section VIII. In Section IX, we investigate an open problem regarding the convergence speed of the error from a viewpoint of approximate recoverability. Section X analyzes the power of a LOCC protocol for transmitting classical information. In Section XI, we provide a lower bound on the cost of resources for an arbitrary LOCC protocol. Conclusions are given in Section XII. The detailed proofs of lemmas and theorems in the main part are presented in Appendices.

*Notations.*  $|\Phi_d\rangle$ ,  $|\Phi_{K_n}\rangle$  and  $|\Phi_{L_n}\rangle$  represent the maximally entangled state with the Schmidt rank  $d$ ,  $K_n$  and  $L_n$ , respectively.  $\pi_d$  is the maximally mixed state of rank  $d$ . The fidelity and the trace distance between two quantum states  $\rho$  and  $\sigma$  are denoted by  $F(\rho, \sigma)$  and  $\|\rho - \sigma\|_1$ , respectively. We abbreviate  $F(\rho, |\psi\rangle\langle\psi|)$  as  $F(\rho, |\psi\rangle)$ . For a quantum operation  $\mathcal{E}$ , we abbreviate  $\mathcal{E}(|\psi\rangle\langle\psi|)$  as  $\mathcal{E}(|\psi\rangle)$ . Otherwise we follow the notations introduced in [12].

## II. FORMULATIONS

Suppose Alice and Bob are given a sequence of bipartite quantum states  $|\psi_{i_1}\rangle^{AB} \dots |\psi_{i_n}\rangle^{AB}$ , generated by an i.i.d. quantum information source of an ensemble  $\{p_i, \psi_i\}_i$ . We assume that the source is completely mixed, in the sense that

$$\sum_i p_i |\psi_i\rangle\langle\psi_i|^{AB} = \pi_d^A \otimes \pi_d^B. \quad (1)$$

Alice and Bob perform the same bipartite unitary  $U^{AB}$  on each of  $|\psi_{i_1}\rangle^{A_1 B_1}, \dots, |\psi_{i_n}\rangle^{A_n B_n}$  by LOCC using a resource state  $\Phi_{K_n}^{A_0 B_0}$ , in such a way that the average error vanishes in the limit of  $n \rightarrow \infty$ . Following the formulation of the Schumacher compression [15], we assume that Alice and Bob do not know the ensemble  $\{p_i, \psi_i\}_i$ , but know that the average state is completely mixed.

Equivalently, we consider a task in which Alice and Bob apply  $(U^{AB})^{\otimes n}$  on  $(|\Phi_d\rangle^{A_0 A_1} |\Phi_d\rangle^{B_0 B_1})^{\otimes n}$  by LOCC using a resource state  $\Phi_{K_n}^{A_0 B_0}$ . Here,  $R_A$  and  $R_B$  are imaginary reference systems that are inaccessible to Alice and Bob

(see Figure 1). Our interest is to find the minimum cost of entanglement, forward and backward classical communication per copy for accomplishing this task. Rigorous definitions are given below.

*Definition 1* Let  $U$  be a bipartite unitary acting on two  $d$ -dimensional quantum systems  $A$  and  $B$ . Let Alice and Bob have quantum registers  $\{A_0, A_1\}$  and  $\{B_0, B_1\}$ , respectively, and let  $\mathcal{M}_n$  be a quantum operation from  $A^n A_0 \otimes B^n B_0$  to  $A^n A_1 \otimes B^n B_1$ .  $\mathcal{M}_n$  is called an  $(r, n, \epsilon)$ -protocol for implementing  $U$  if  $\mathcal{M}_n$  is an  $r$ -round LOCC that satisfies

$$F(\rho(\mathcal{M}_n), |\Psi_U\rangle^{\otimes n} |\Phi_{L_n}\rangle^{A_1 B_1}) \geq 1 - \epsilon, \quad (2)$$

where

$$|\Psi_U\rangle := U^{AB} |\Phi_d\rangle^{A_0 A_1} |\Phi_d\rangle^{B_0 B_1}$$

and

$$\rho(\mathcal{M}_n) := \mathcal{M}_n(|\Phi_d^{A_0 A_1}\rangle^{\otimes n} |\Phi_d^{B_0 B_1}\rangle^{\otimes n} |\Phi_{K_n}\rangle^{A_0 B_0}). \quad (3)$$

The entanglement cost of  $\mathcal{M}_n$  is defined by  $\log K_n - \log L_n$ . The forward and backward classical communication cost of  $\mathcal{M}_n$  are defined as the sum of numbers of classical bits transmitted from Alice to Bob and from Bob to Alice, respectively, in  $\mathcal{M}_n$ .

*Definition 2* A rate triplet  $(E, C_f, C_b)$  is said to be achievable by an  $r$ -round protocol if, for any  $\epsilon > 0$ , there exists  $n_\epsilon$  such that for any  $n \geq n_\epsilon$ , we find an  $(r, n, \epsilon)$ -protocol for implementing  $U$  with the entanglement cost  $nE$ , forward classical communication cost  $nC_f$  and backward classical communication cost  $nC_b$ .

Condition (2) implies that, for *almost all* input states  $|\psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$ , the final state  $\mathcal{M}_n(|\psi\rangle^{\bar{A}\bar{B}} |\Phi_{K_n}\rangle^{A_0 B_0})$  is close to the desired state  $(U^{\otimes n} |\psi\rangle)^{\bar{A}\bar{B}} |\Phi_{L_n}\rangle^{A_0 B_0}$ . That is, we have

$$\int_{\text{Haar}} p(d\psi) F(\rho(\mathcal{M}_n, \psi), (U^{\otimes n} |\psi\rangle)^{\bar{A}\bar{B}} |\Phi_{L_n}\rangle^{A_0 B_0}) \geq 1 - \epsilon \quad (4)$$

for

$$\rho(\mathcal{M}_n, \psi) := \mathcal{M}_n(|\psi\rangle^{\bar{A}\bar{B}} |\Phi_{K_n}\rangle^{A_0 B_0}),$$

where the average is taken with respect to the Haar measure on  $\mathcal{H}^A \otimes \mathcal{H}^B$  (see Appendix B-A for a proof).

Ideally, we should relax Condition (1), and require that the protocol be *universal* in the sense that  $\mathcal{M}_n(|\psi\rangle^{\bar{A}\bar{B}})$  is close to the desired state  $U^{\otimes n} |\psi\rangle$  for *all* input states  $|\psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$ . In this paper, however, we require Condition (1) for simplicity.

## III. RESULTS

The results of this paper are summarized as follows.

### A. Result 1: Achievable Rate Region for Two-Round Protocols

In this paper, we mainly consider two-round protocols (i.e.  $r = 2$ ) starting with Alice's operation. In general, such a protocol proceeds as follows: Alice first performs a measurement and communicates the outcome to Bob; Bob then performs a measurement and communicates the outcome to Alice; and, finally, Alice performs an operation. The first main result of this paper is that the optimal rate of the cost of entanglement and of classical communication in a two-round protocol are given by a parameter called the *Markovianizing cost of the unitary*, under an additional requirement that the error vanishes faster than  $1/n^4$ .

To present a rigorous statement, let us introduce a concept of the Markovianizing cost of tripartite quantum states [12]. A tripartite quantum state  $\Upsilon^{ABC}$  is called a *Markov state conditioned by B* if it satisfies  $I(A : C|B)_\Upsilon = 0$  [16]. *Markovianization* is a task in which  $n$  copies of a tripartite state  $\rho^{ABC}$  is transformed by a randomizing operation on  $A^n$  to a Markov state conditioned by  $B^n$ . The *Markovianizing cost of  $\rho^{ABC}$*  is defined as the minimum cost of randomness per copy required for the task, in an asymptotic limit of infinite copies and vanishingly small error. A rigorous definition is as follows.

**Definition 3** A tripartite state  $\rho^{ABC}$  is *Markovianized* with the randomness cost  $R$  on  $A$ , conditioned by  $B$ , if the following statement holds. That is, for any  $\epsilon > 0$ , there exists  $n_\epsilon$  such that for any  $n \geq n_\epsilon$ , we find a random unitary operation  $\mathcal{V}_n : \tau \mapsto 2^{-nR} \sum_{k=1}^{2^{nR}} V_k \tau V_k^\dagger$  on  $A^n$  and a Markov state  $\Upsilon^{A^n B^n C^n}$  conditioned by  $B^n$  that satisfy

$$\left\| \mathcal{V}_n^{A^n}(\rho^{\otimes n}) - \Upsilon^{A^n B^n C^n} \right\|_1 \leq \epsilon. \quad (5)$$

The *Markovianizing cost of  $\rho^{ABC}$*  is defined as  $M_{A|B}(\rho^{ABC}) := \inf\{R \mid \rho^{ABC} \text{ is Markovianized with the randomness cost } R \text{ on } A, \text{ conditioned by } B\}$ .

We extend the notion of Markovianizing cost to a bipartite unitary as follows.

**Definition 4** Let  $U$  be a bipartite unitary acting on two  $d$ -level systems  $A$  and  $B$ , and consider a “tripartite” state

$$|\Psi_U\rangle^{AR_A(BR_B)} := (U^{AB} \otimes I^{R_A R_B})|\Phi_d\rangle^{AR_A}|\Phi_d\rangle^{BR_B} \quad (6)$$

by regarding  $B$  and  $R_B$  as a single system. The Markovianizing cost of  $U$  is defined as  $M(U) := M_{A|R_A}(\Psi_U^{AR_A(BR_B)})$ .

The main result of this paper is presented by the following theorem. The proofs are given in Section VI and the corresponding appendices, after preparatory arguments in Section IV and V.

#### Theorem 5

- *Direct:* A rate triplet  $(E, C_f, C_b)$  is achievable by a two-round protocol for implementing  $U$  if  $E, C_f, C_b > M(U^\dagger)$ .

- *Converse:* A rate triplet  $(E, C_f, C_b)$  is achievable by a two-round protocol for implementing  $U$  only if  $E, C_f, C_b \geq M(U^\dagger)$ , if we additionally require in Definition 2 that

$$\lim_{\epsilon \rightarrow 0} \epsilon \cdot n_\epsilon^4 = 0. \quad (7)$$

It is left open whether the same converse bound holds when we drop Condition (7). As we will discuss in Section IX in detail, this question is directly related to another question of whether Equality (14) holds without Condition (7). At the core of these questions lies an open problem regarding an “asymptotic symmetry” of approximate recoverability.

### B. Result 2: General Lower Bound on the Cost of Resources

Any bipartite unitary  $U$  on  $AB$  is decomposed as

$$U^{AB} = \sum_{s=0}^{d^2-1} c_s E_s^A \otimes F_s^B, \quad (8)$$

where  $c_s$  ( $s = 0, \dots, d^2 - 1$ ) are nonnegative real numbers that satisfy

$$\sum_{s=0}^{d^2-1} c_s^2 = 1, \quad c_s \geq 0 \quad (\forall s),$$

and  $E_s \in \mathcal{L}(\mathcal{H}^A)$ ,  $F_s \in \mathcal{L}(\mathcal{H}^B)$  are linear operators which are orthonormal with respect to the Hilbert-Schmidt inner product, i.e.,

$$\frac{1}{d} \text{Tr}[E_s^\dagger E_{s'}] = \frac{1}{d} \text{Tr}[F_s^\dagger F_{s'}] = \delta_{ss'}. \quad (9)$$

The Shannon entropy of  $\{c_s^2\}_s$  is called the *Schmidt strength* of  $U$ . We denote it by  $K(U)$ , that is,

$$K(U) := H(\{c_s^2\}_s) = - \sum_s c_s^2 \log c_s^2. \quad (10)$$

The following theorem states that a lower bound on the minimum cost of entanglement and classical communication, in a protocol with arbitrary number of rounds of communication, is given by the Schmidt strength of the unitary. Proofs are given in Section XI and the corresponding appendices.

**Theorem 6** A rate triplet  $(E, C_f, C_b)$  is achievable only if  $E, C_f, C_b \geq K(U)$ .

## IV. PRELIMINARIES

We review an alternative definition of the Markovianizing cost [13], [14], in addition to state merging [17], [18]. The results reviewed here are used in the following sections to prove Theorem 5.

### A. Markovianization in terms of Recoverability

It is proved in [16] that the following conditions are equivalent:

- 1) *Vanishing QCMDI*:  $\rho^{ABC}$  is a Markov state conditioned by  $B$ , i.e., it satisfies

$$I(A : C|B)_\rho = 0.$$

- 2) *Recoverability*:  $\rho^{ABC}$  is recoverable from its bipartite reduced state on  $AB$  and  $BC$ , that is, there exist quantum operations  $\mathcal{R} : B \rightarrow AB$  and  $\mathcal{R}' : B \rightarrow BC$  such that

$$\rho^{ABC} = \mathcal{R}(\rho^{BC}) = \mathcal{R}'(\rho^{AB}). \quad (11)$$

Based on this fact, the *Markovianizing cost in terms of recoverability* is introduced in [14]. In the same way as Definition 3, we consider a task in which  $n$  copies of a tripartite quantum state  $\rho^{ABC}$  is transformed by a random unitary operation on  $A^n$ . Instead of requiring that the state after the operation satisfies Condition (5), however, we now require that the state satisfies Condition (11) up to a small error  $\epsilon$ . Rigorous definitions are as follows.

**Definition 7** A tripartite state  $\rho^{ABC}$  is said to be  $\epsilon$ -recoverable from  $BC$  if there exists a quantum operation  $\mathcal{R} : B \rightarrow AB$  such that

$$\|\rho^{ABC} - \mathcal{R}(\rho^{BC})\|_1 \leq \epsilon.$$

$\rho^{ABC}$  is  $\epsilon$ -recoverable from  $AB$  if there exists a quantum operation  $\mathcal{R}' : B \rightarrow AB$  such that

$$\|\rho^{ABC} - \mathcal{R}'(\rho^{AB})\|_1 \leq \epsilon.$$

**Definition 8** A tripartite state  $\rho^{ABC}$  is *Markovianized with the randomness cost  $R$  on  $A$ , in terms of recoverability from  $BC$* , if it holds that: for any  $\epsilon > 0$ , there exists  $n_\epsilon$  such that for any  $n \geq n_\epsilon$ , we find a random unitary operation

$$\mathcal{V}_n : \tau \mapsto 2^{-nR} \sum_{k=1}^{2^{nR}} V_k \tau V_k^\dagger$$

on  $A^n$ , so that  $\mathcal{V}_n((\rho^{ABC})^{\otimes n})$  is  $\epsilon$ -recoverable from  $B^n C^n$ .

This allows us to define the *Markovianizing cost of  $\rho^{ABC}$  in terms of recoverability from  $BC$*  as  $M_{A|BC}^R(\rho^{ABC}) := \inf\{R \mid \rho^{ABC} \text{ is Markovianized with the randomness cost } R \text{ on } A, \text{ in terms of recoverability from } BC\}$ .

We also consider a Markovianization induced by a measurement, supplemented by auxiliary entanglement resource.

**Definition 9** Consider a tripartite pure state  $|\Psi\rangle^{ABC}$ , and let  $A_0$  and  $G$  be additional quantum systems. Suppose that for any  $\epsilon > 0$ , there exists  $n_\epsilon$  such that for any  $n \geq n_\epsilon$ , we find a pure state  $|\varrho_n\rangle^{A_0 G}$  and a measurement on  $\bar{A}A_0$ , which is described by a set of measurement operators  $\{M_k^{\bar{A}A_0 \rightarrow A'}\}_{k \in \mathbb{K}}$ , satisfying the following conditions:

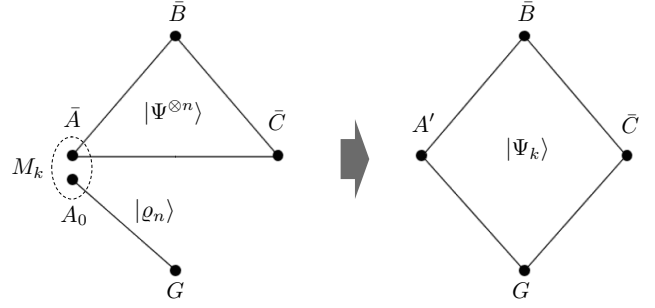


Fig. 2. A graphical representation of Markovianization of a pure state by a measurement with an auxiliary entangled resource. After the measurement, the reduced state on  $A'\bar{B}\bar{C}$  should be an approximately recoverable state.

- 1) The measurement does not significantly change the reduced state on  $\bar{B}\bar{C}$  on average, i.e.,

$$\sum_{k \in \mathbb{K}} p_k \left\| (\Psi^{\otimes n})^{\bar{B}\bar{C}} - \Psi_k^{\bar{B}\bar{C}} \right\|_1 \leq \epsilon, \quad (12)$$

where  $p_k$  is the probability of obtaining the outcome  $k$ , and  $\Psi_k$  is the post-measurement state corresponding to the outcome  $k$ .

- 2) The post-measurement state is approximately recoverable on average, that is, there exist linear CPTP maps  $\mathcal{R}_{n,k} : \bar{B} \rightarrow \bar{B}\bar{C}$  ( $k \in \mathbb{K}$ ) satisfying

$$\sum_{k \in \mathbb{K}} p_k \left\| \Psi_k^{A'\bar{B}\bar{C}} - \mathcal{R}_{n,k}(\Psi_k^{A'\bar{B}}) \right\|_1 \leq \epsilon. \quad (13)$$

- 3) The correlation between  $\bar{B}\bar{C}$  and  $G$  produced by the measurement is at most  $nR$  bits in QMI, that is,

$$I(\bar{B}\bar{C} : G)_{av} := \sum_{k \in \mathbb{K}} p_k I(\bar{B}\bar{C} : G)_{\Psi_k} \leq nR.$$

Then  $|\Psi\rangle^{ABC}$  is said to be *Markovianized with the correlation production  $R$  by a measurement on  $A$ , in terms of recoverability from  $AB$* .

Correspondingly, the *measurement-induced Markovianizing cost of  $|\Psi\rangle^{ABC}$  in terms of recoverability from  $AB$*  is defined as  $M_{A|AB}^{R,m}(\Psi^{ABC}) := \inf\{R \mid |\Psi\rangle^{ABC} \text{ is Markovianized with the correlation production } R \text{ by a measurement on } A, \text{ in terms of recoverability from } AB\}$ .

The two types of Markovianizing costs defined above are equal to that in Definition 3 for pure states, if we impose an additional requirement on the convergence speed of the error in Definition 9 [14].

**Theorem 10** (Theorem 11 and 14 in [14]) For any tripartite pure state  $|\Psi\rangle^{ABC}$ , we have

$$M_{A|B}(\Psi^{ABC}) = M_{A|BC}^R(\Psi^{ABC}) = M_{A|AB}^{R,m}(\Psi^{ABC}), \quad (14)$$

if we additionally require in Definition 9 that

$$\lim_{\epsilon \rightarrow 0} \epsilon \cdot n_\epsilon = 0. \quad (15)$$

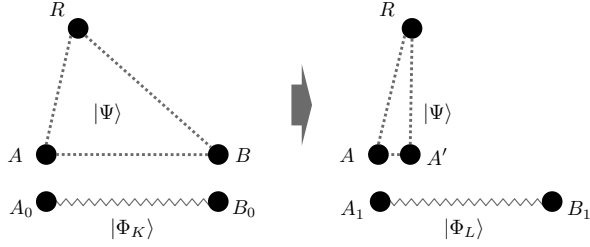


Fig. 3. State merging is a task in which Bob transfers his share of  $|\Psi\rangle_{ABR}$  to Alice.  $R$  is an inaccessible reference system. For the sake of presentation, we consider Bob as the sender and Alice as the receiver.

The following lemma relates the Markovianizing cost to other entropic quantities characterizing the state transformation induced by a Markovianizing measurement.

*Lemma 11* (See Appendix B-B for a proof.) There exists a nonnegative function  $\tilde{\xi}(\delta)$ , such that  $\lim_{\delta \rightarrow 0} \tilde{\xi}(\delta) = 0$ , and for any  $n \in \mathbb{N}$ ,  $\epsilon > 0$ ,  $\delta \in (0, 1]$  and  $\{M_k^{AA_0 \rightarrow A'}\}_k$  satisfying  $\epsilon \cdot n \leq \delta$  and Inequalities (12) and (13), we have

$$\begin{aligned} H(\{p_k\}_{k \in \mathbb{K}}) &\geq \Delta S(A')_{av} \geq \Delta S(A')_{av} - \Delta S(G)_{av} \\ &\geq nM_{A|B}(\Psi^{ABC}) - n\tilde{\xi}(\delta), \end{aligned} \quad (16)$$

where

$$\begin{aligned} \Delta S(A')_{av} &:= nS(A) + S(A_0)_{\varrho_n} - \sum_{k \in \mathbb{K}} S(A')_{\Psi_k}, \\ \Delta S(G)_{av} &:= S(G)_{\varrho_n} - \sum_{k \in \mathbb{K}} S(G)_{\Psi_k}. \end{aligned}$$

*Remark.* It has been left open whether Equality (14) holds when we remove Condition (15). See Conjecture 13 of [14].

### B. State Merging

Suppose Alice and Bob share a tripartite pure state  $|\Psi\rangle_{ABR}$  with an inaccessible reference system  $R$ . State merging ([17], [18]) is a task in which Bob sends his share of  $\Psi$  to Alice so that Alice has both  $A$  and  $B$  parts of  $\Psi$ , or equivalently, so that Alice has the whole part of the purification of  $\Psi^R$ . (See Figure 3. For later convenience, we exchange roles of Alice and Bob in the standard formulation.) Our concern is the cost of entanglement and classical communication required for state merging. A rigorous definition is given as follows.

*Definition 12* Consider a tripartite pure state  $|\Psi\rangle_{ABR}$ . Let Alice and Bob have quantum systems  $\{A_B, A_0, A_1\}$  and  $\{B_0, B_1\}$ , respectively, where  $A_B$  is assumed to be identical to  $B$ . The following protocol  $\mathcal{N}$  consisting of a sequence of quantum operations is called state merging of  $\Psi$  with error  $\epsilon$ , entanglement cost  $\log K - \log L$  and classical communication cost  $C$ . Here,  $\mathcal{N} : AA_0BB_0 \rightarrow AA_BA_1B_1$  is a one-way LOCC from Bob to Alice, such that

$$F(\rho(\mathcal{N}), |\Psi\rangle^{AA_B R} |\Phi_L\rangle^{A_1 B_1}) \geq 1 - \epsilon \quad (17)$$

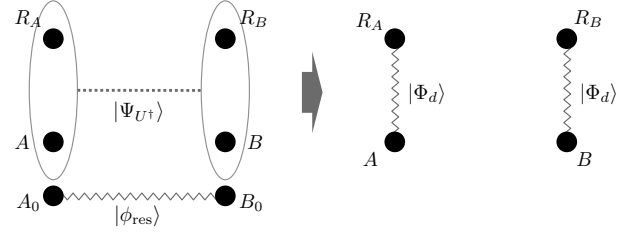


Fig. 4. A graphical representation of a task that we analyze in Section V. The task is to destroy correlation between  $AR_A$  and  $BR_B$ , while preserving the maximal entanglement between  $AB$  and  $R_A R_B$ . A certain amount of shared entanglement must be consumed to accomplish this task, since reference systems  $R_A$  and  $R_B$  are inaccessible.

for

$$\rho(\mathcal{N}) := \mathcal{N}(|\Psi\rangle^{ABR} |\Phi_K\rangle^{A_0 B_0}),$$

$K$  and  $L$  are natural numbers, and  $\Phi_K$  and  $\Phi_L$  are the maximally entangled states with the Schmidt rank  $K$  and  $L$ , respectively.  $C$  is the total amount of classical communication transmitted from Bob to Alice in  $\mathcal{M}$ , measured in bits.

There always exists a state merging with an error determined by the initial state. The following theorem is obtained as a corollary of Proposition 3 and 4 in [18] by letting  $L = 1$ .

*Theorem 13* Let  $D_A := (\text{Tr}[(\Psi^A)^2])^{-1}$  and  $r_B := \text{rank}[\Psi^B]$ . There exists a state merging of  $\Psi$  with entanglement cost 0, classical communication cost  $C = \log r_B$  and error

$$\epsilon \leq 2\sqrt{2}(\sqrt{d_R/D_A} + 1/r_B)^{1/2}. \quad (18)$$

It is also proved in [18] that the cost of entanglement and classical communication in a state merging are bounded from below as

$$\log K - \log L \gtrsim nS(B|A)_\Psi, \quad C \gtrsim nI(B : R)_\Psi, \quad (19)$$

when  $\epsilon$  is sufficiently small. See Appendix B-C for details.

### V. SINGLE-SHOT TWO-ROUND PROTOCOLS

In this section, we consider  $n = 1$  (single-shot) case, and analyze a single-shot protocol  $\mathcal{M}$  for implementing  $U$  by two-round LOCC assisted by shared entanglement. The results obtained here are then applied to the asymptotic situation in Section VI.

Let  $\mathcal{M} : AA_0 \otimes BB_0 \rightarrow AA_1 \otimes BB_1$  be a two-round LOCC protocol for implementing  $U$ .  $\mathcal{M}$  succeeds in implementing  $U$  with high fidelity, if

$$F(\rho(\mathcal{M})^{AR_A BR_B}, |\Psi_U\rangle) \geq 1 - \epsilon \quad (20)$$

for some small  $\epsilon$ , where

$$\rho(\mathcal{M}) := \mathcal{M}(|\Phi_d\rangle^{AR_A} |\Phi_d\rangle^{BR_B} |\phi_{\text{res}}\rangle^{A_0 B_0})$$

and  $\phi_{\text{res}}$  is a pure resource state shared in advance. Since we have  $\Phi_d^A \otimes \Phi_d^B = \Psi_U^{AB}$ , and all purifications are equivalent up

to local unitary transformations, there exists a unitary  $\hat{U}$  on  $R_A R_B$  such that

$$|\Phi_d\rangle^{AR_A} |\Phi_d\rangle^{BR_B} = \hat{U}^{R_A R_B} |\Psi_U\rangle^{AR_A BR_B}.$$

Applying  $U^{\dagger AB}$  on both sides yields

$$\begin{aligned} |\Psi_{U^\dagger}\rangle^{AR_A BR_B} &:= U^{\dagger AB} |\Phi_d\rangle^{AR_A} |\Phi_d\rangle^{BR_B} \\ &= \hat{U}^{R_A R_B} |\Phi_d\rangle^{AR_A} |\Phi_d\rangle^{BR_B}, \end{aligned} \quad (21)$$

which leads to

$$\begin{aligned} \rho(\mathcal{M}, U^\dagger) &:= \mathcal{M}(|\Psi_{U^\dagger}\rangle^{AR_A BR_B} |\phi_{\text{res}}\rangle^{A_0 B_0}) \\ &= \hat{U}^{R_A R_B} \rho(\mathcal{M})^{AR_A BR_B} \hat{U}^{\dagger R_A R_B}. \end{aligned}$$

Note that  $\mathcal{M}$  does not act on  $R_A R_B$ . Therefore, due to the unitary invariance of the fidelity, Condition (20) is equivalent to

$$F(\rho(\mathcal{M}, U^\dagger)^{ABR_A R_B}, |\Phi_d\rangle^{AR_A} |\Phi_d\rangle^{BR_B}) \geq 1 - \epsilon. \quad (22)$$

While  $|\Phi_d\rangle |\Phi_d\rangle$  obviously has no correlation between  $AR_A$  and  $BR_B$ ,  $|\Psi_{U^\dagger}\rangle$  has a certain amount of entanglement depending on  $U^\dagger$ . Thus, for a given initial state  $|\Psi_{U^\dagger}\rangle$  and a resource state  $|\phi_{\text{res}}\rangle$ , a successful protocol  $\mathcal{M}$  decouples  $AR_A$  and  $BR_B$  while preserving the maximal entanglement between  $AB$  and  $R_A R_B$  (Figure 4). Observe that both  $|\Psi_{U^\dagger}\rangle$  and  $|\Phi_d\rangle |\Phi_d\rangle$  are maximally entangled states between  $AB$  and  $R_A R_B$  with Schmidt rank  $d^2$ .

The main goal of this section is to derive conditions on operations that comprise  $\mathcal{M}$ , for the protocol to succeed with high fidelity. It turns out that any successful protocol can be described as a combination of Markovianization and a subsequent state merging. Consequently, as we describe in detail in Section VI, the minimum cost of resources is derived by combining results on Markovianization and state merging presented in Section IV.

In the following, we fix a unitary  $U$  acting on  $AB$ , and denote  $\Psi_{U^\dagger}$  simply by  $\Psi$ . Without loss of generality, we assume that the two-round protocol  $\mathcal{M}$  proceeds as follows (Figure 5):

1. Alice performs a measurement on  $AA_0$ , which is described by a set of measurement operators  $\mathbb{M} = \{M_k^{AA_0 \rightarrow A'}\}_k$ , and obtains an outcome  $k$ .
2. Alice communicates  $k$  to Bob.
3. Bob performs a measurement on  $BB_0$ , described by  $\mathbb{N}_k = \{N_{l|k}^{BB_0 \rightarrow BB_1}\}_l$ , and obtains an outcome  $l$ .
4. Bob communicates  $l$  to Alice.
5. Alice performs an operation which is described by a linear CPTP map  $\mathcal{O}_{kl} : A' \rightarrow AA_1$ .

Here,  $A'$  is the output system of Alice's measurement such that

$$\dim \mathcal{H}^{A'} \leq \dim \mathcal{H}^A \times \dim \mathcal{H}^{A_0}. \quad (23)$$

We denote the set of outcomes of Alice's measurement by  $\mathbb{K}$ . (See Remark at the end of Appendix D-A for a treatment of protocols in which not all information about the measurement outcome  $k$  is communicated to Bob.)

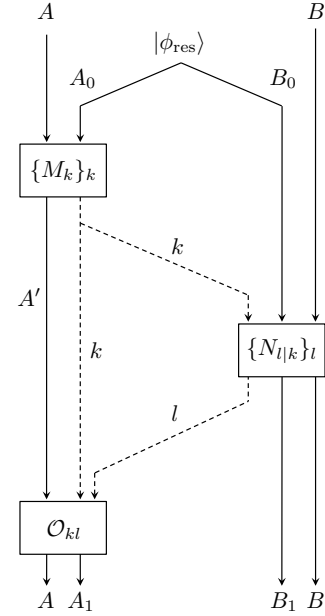


Fig. 5. A graphical description of how the two-round protocol  $\mathcal{M}$  proceeds. The black lines represent quantum systems, the two dotted lines indicate classical communication, and the three boxes represent quantum measurements and operations.

#### A. Conditions on Alice's Measurement

Let us first discuss general conditions regarding state transformations by Alice's measurement. For a fixed  $\phi_{\text{res}}$ , and for any linear operator  $M : \mathcal{H}^A \otimes \mathcal{H}^{A_0} \rightarrow \mathcal{H}^{A'}$  such that  $M^\dagger M \leq I$ , we define a map  $\mathcal{E}_M$  by

$$\begin{aligned} \mathcal{E}_M(\tau) &:= p_M^{-1} M(\tau^A \otimes \phi_{\text{res}}^{A_0}) M^\dagger, \\ p_M &:= \text{Tr}[M(\tau^A \otimes \phi_{\text{res}}^{A_0}) M^\dagger]. \end{aligned}$$

We call  $\mathcal{E}_M$  as an  $M$ -induced map.  $M$  is supposed to be an element of  $\mathbb{M}$ , in which case  $p_M$  describes the probability of obtaining a certain measurement outcome corresponding to  $M$ . Note that  $p_M$  depends on the input state  $\tau$  in general. Consequently, the  $M$ -induced map is not necessarily a linear map. The linearity of  $\mathcal{E}_M$  is equivalent to the independence of  $p_M$  from  $\tau$ , which indicates that the measurement is *oblivious* to the input, in the sense that it does not extract any information about the input state. This obliviousness condition plays an important role in proofs of most of the lemmas in this section (as well as in [10]). Thus we introduce an equivalent definition of approximate obliviousness as follows.

**Definition 14** An  $M$ -induced map is  $\varsigma$ -oblivious if it satisfies

$$\left\| \Phi_M^{RA} - \pi_d^{RA} \right\|_1 \leq \varsigma,$$

where  $\Phi_M^{A'R_A} := \mathcal{E}_M^A(\Phi_d^{AR_A})$ . A measurement  $\mathbb{M}$  is  $\varsigma$ -oblivious if an  $M_k$ -induced map is  $\varsigma_k$ -oblivious for each  $k$  and  $\sum_{k \in \mathbb{K}} p_k \varsigma_k \leq \varsigma$ .

We introduce some other conditions on Alice's measurement. In the following, we denote  $\mathcal{E}_M(\Psi)$  as  $\Psi_M$ .

*Definition 15* An  $M$ -induced map is  $\mu$ -decoupling between  $A'R_A$  and  $R_B$  if it satisfies

$$\left\| \Psi_M^{A'R_A R_B} - \Psi_M^{A'R_A} \otimes \Psi_M^{R_B} \right\|_1 \leq \mu. \quad (24)$$

A measurement  $\mathbb{M}$  is  $\mu$ -decoupling between  $A'R_A$  and  $R_B$  if an  $M_k$ -induced map is  $\mu_k$ -decoupling between  $A'R_A$  and  $R_B$  for each  $k$  and  $\sum_{k \in \mathbb{K}} p_k \mu_k \leq \mu$ .

As in Definition 4, we now consider  $B$  and  $R_B$  as a single system and regard  $\Psi_M^{A'R_A B R_B}$  as a “tripartite” state on  $A'$ ,  $R_A$  and  $B R_B$ .

*Definition 16* An  $M$ -induced map is  $\nu$ -Markovianizing from  $R_A B R_B$  if  $\Psi_M^{A'R_A B R_B}$  is  $\nu$ -recoverable from  $R_A B R_B$ , that is, if there exists a linear CPTP map  $\mathcal{R} : R_A \rightarrow A'R_A$  such that

$$\left\| \Psi_M^{A'R_A B R_B} - \mathcal{R}(\Psi_M^{R_A B R_B}) \right\|_1 \leq \nu.$$

A measurement  $\mathbb{M}$  is  $\nu$ -Markovianizing from  $R_A B R_B$  if an  $M_k$ -induced map is  $\nu_k$ -Markovianizing from  $R_A B R_B$  for each  $k$  and  $\sum_{k \in \mathbb{K}} p_k \nu_k \leq \nu$ .

*Definition 17* An  $M$ -induced map is  $\iota$ -Markovianizing from  $A'R_A$  if  $\Psi_M^{A'R_A B R_B}$  is  $\iota$ -recoverable from  $A'R_A$ , that is, if there exists a linear CPTP map  $\mathcal{R}' : R_A \rightarrow R_A B R_B$  such that

$$\left\| \Psi_M^{A'R_A B R_B} - \mathcal{R}'(\Psi_M^{A'R_A}) \right\|_1 \leq \iota.$$

A measurement  $\mathbb{M}$  is  $\iota$ -Markovianizing from  $A'R_A$  if an  $M_k$ -induced map is  $\iota_k$ -Markovianizing from  $A'R_A$  for each  $k$  and  $\sum_{k \in \mathbb{K}} p_k \iota_k \leq \iota$ .

The following two lemmas are at the core of the proofs of the main result, which translates the problem of finding the optimal costs for implementing a bipartite unitary to that of computing the Markovianizing cost of a particular state.

*Lemma 18* A measurement  $\mathbb{M}$  is  $(3\varsigma + 2\nu)$ -decoupling between  $A'R_A$  and  $R_B$  if it is  $\varsigma$ -oblivious and  $\nu$ -Markovianizing from  $R_A B R_B$ .

*Lemma 19* A measurement  $\mathbb{M}$  is  $(\varsigma + \mu)$ -Markovianizing from  $A'R_A$  if it is  $\varsigma$ -oblivious and  $\mu$ -decoupling between  $A'R_A$  and  $R_B$ .

Let us describe a simplified version of the proof of the above two lemmas in the case of  $\mu = \nu = \varsigma = 0$ . The conditions of  $\epsilon$ -Markovianizing in Definition 16 and 17 are then equivalent to the condition that  $\Psi_M^{A'R_A B R_B}$  is a Markov state conditioned by  $R_A$ . Suppose an  $M$ -induced map is 0-oblivious, which implies  $\Phi_M^{R_A} = \pi_d^{R_A}$ . Using (21), we see that

$$\Psi_M^{A'R_A B R_B} = \hat{U}^{R_A R_B} (\Phi_M^{A'R_A} \otimes \Phi_d^{B R_B}) \hat{U}^{\dagger R_A R_B}, \quad (25)$$

and consequently,

$$\begin{aligned} \Psi_M^{R_A R_B} &= \hat{U}^{R_A R_B} (\pi_d^{R_A} \otimes \pi_d^{R_B}) \hat{U}^{\dagger R_A R_B} \\ &= \pi_d^{R_A} \otimes \pi_d^{R_B}. \end{aligned} \quad (26)$$

Therefore, for the state  $\Psi_M^{A'R_A B R_B}$ , we have  $I(A' : B | R_A R_B) = 0$  due to the local unitary invariance of QCMI, as well as  $I(R_A : R_B) = 0$ . It follows that

$$\begin{aligned} I(A' : B R_B | R_A) &= I(A' : R_B | R_A) + I(A' : B | R_A R_B) \\ &= I(A' R_A : R_B) - I(R_A : R_B) \\ &= I(A' R_A : R_B), \end{aligned}$$

which implies the equivalence between the conditions of decoupling and Markovianizing under the condition of obliviousness.

For rigorous proofs, we need to relax the “exact” condition ( $\mu = \nu = \varsigma = 0$ ) to the “approximate” condition ( $\mu, \nu, \varsigma > 0$ ). See Appendices C-A and C-B for details.

### B. Conditions for Achievability

For the proof of the direct part of Theorem 5, let us consider how to construct a successful protocol. Let  $\tilde{B}$  be a register on Bob’s side which has a sufficiently large dimension. The following lemma states that Markovianization by Alice’s measurement is a sufficient condition for the success of the first half of  $\mathcal{M}$ , in which  $\Phi_d^{B R_B}$  is obtained from  $|\Psi\rangle$ .

*Lemma 20* (See Appendix C-C for a proof.) Suppose that a measurement  $\mathbb{M}$  is 0-oblivious and  $\mu$ -decoupling between  $A'R_A$  and  $R_B$ ,  $\mu \in (0, 1]$ , and that  $\Psi_{M_k}^{A'R_A R_B}$  does not depend on  $k$ . Then, there exist pure states  $|\tilde{\Psi}^p\rangle^{A'R_A \tilde{B}}$ ,  $|\Psi'\rangle^{A'R_A \tilde{B} B R_B}$  and isometries  $W_k^{B B_0 \rightarrow B \tilde{B}}$  ( $k \in \mathbb{K}$ ) such that

$$\left\| |\Psi'\rangle\langle\Psi'| - (\tilde{\Psi}^p)^{A'R_A \tilde{B}} \otimes \Phi_d^{B R_B} \right\|_1 \leq 5\sqrt[4]{\mu} \quad (27)$$

and  $|\Psi'\rangle = W_k |\Psi_{M_k}\rangle$  for any  $k \in \mathbb{K}$ . In addition,  $|\tilde{\Psi}^p\rangle$  satisfies

$$\left\| (\tilde{\Psi}^p)^{R_A} - \pi_d^{R_A} \right\|_1 \leq 3\sqrt[4]{\mu}, \quad \text{rank}[(\tilde{\Psi}^p)^{\tilde{B}}] \leq \dim \mathcal{H}^{B_0}.$$

The following lemma immediately follows from Lemma 18 and 20.

*Lemma 21* Suppose that a measurement  $\mathbb{M}$  is 0-oblivious and  $\nu$ -Markovianizing from  $R_A B R_B$ ,  $\nu \in (0, 1/2]$ , and that  $\Psi_{M_k}^{A'R_A R_B}$  does not depend on  $k$ . Then, there exist pure states  $|\tilde{\Psi}^p\rangle^{A'R_A \tilde{B}}$ ,  $|\Psi'\rangle^{A'R_A \tilde{B} B R_B}$  and isometries  $W_k^{B B_0 \rightarrow B \tilde{B}}$  ( $k \in \mathbb{K}$ ) that satisfy

$$\begin{aligned} \left\| |\Psi'\rangle\langle\Psi'| - (\tilde{\Psi}^p)^{A'R_A \tilde{B}} \otimes \Phi_d^{B R_B} \right\|_1 &\leq 5\sqrt[4]{2\nu}, \\ |\Psi'\rangle &= W_k |\Psi_{M_k}\rangle \quad (\forall k \in \mathbb{K}) \end{aligned}$$

and

$$\left\| (\tilde{\Psi}^p)^{R_A} - \pi_d^{R_A} \right\|_1 \leq 3\sqrt[4]{2\nu}, \quad \text{rank}[(\tilde{\Psi}^p)^{\tilde{B}}] \leq \dim \mathcal{H}^{B_0}. \quad (28)$$

The task remaining after obtaining  $\Phi_d^{B R_B}$  is to obtain  $\Phi_d^{A R_A}$  from tripartite pure states  $|\tilde{\Psi}^p\rangle^{A'R_A \tilde{B}}$ , which is equivalent to performing state merging from Bob to Alice. Consequently, we can construct a successful protocol by combining Markovianization of  $|\Psi\rangle$  and the subsequent state merging of  $|\tilde{\Psi}^p\rangle$  from Bob to Alice.

### C. Conditions for Optimality

For the proof of the converse part of Theorem 5, let us analyze conditions on Alice's measurement imposed by (22). Let  $\mathbb{M} = \{M_k^{AA_0 \rightarrow A'}\}_{k \in \mathbb{K}}$  be Alice's measurement in protocol  $\mathcal{M}$  that satisfies (22). First, conservation of the maximal entanglement between systems  $AB$  and  $R_A R_B$  immediately implies that Alice's measurement must be oblivious. Second, since the final state is close to  $|\Phi_d\rangle^{AR_A} |\Phi_d\rangle^{BR_B}$ , correlation between  $AR_A$  and  $R_B$  is destroyed by  $\mathcal{M}$ . This part of decoupling must be accomplished by Alice's measurement alone, which implies that Alice's measurement must be Markovianizing due to Lemma 19. Hence we obtain the following lemma.

**Lemma 22** (Appendices D-A and D-B) The measurement  $\mathbb{M}$  is  $4\sqrt[4]{\epsilon}$ -oblivious and  $12\sqrt[4]{\epsilon}$ -Markovianizing from  $A'R_A$ .

Let us continue to analyze conditions on Bob's measurement imposed by (22). Let  $B_E$  be an ancillary system, and let  $W_k : BB_0 \rightarrow BB_1 B_E$  be an isometry such that the Naimark extension of Bob's measurement  $\mathbb{N}_k$  is given by  $N_{l|k} = \langle l|^{B_E} W_k$ . Define  $|\Psi'_k\rangle := W_k |\Psi_k\rangle$ . The following lemma states that Bob's measurement is decomposed into two parts: (i) performing an isometry to obtain  $\Phi_d^{BR_B}$ , and (ii) performing a measurement on his share of a "tripartite" pure state on  $A'$ ,  $R_A$  and  $B_1 B_E$ .

**Lemma 23** (Appendix D-C) There exist pure states  $|\Psi_k^p\rangle^{A'R_A B_1 B_E}$  ( $k \in \mathbb{K}$ ) such that

$$\sum_{k \in \mathbb{K}} p_k \left\| \Psi'_k - (\Psi_k^p)^{A'R_A B_1 B_E} \otimes \Phi_d^{BR_B} \right\|_1 \leq 4\sqrt[4]{\epsilon}.$$

By the measurement on  $B_E$  described by  $\{|l\rangle\langle l|\}_l$  and an operation on  $A'$  depending on  $k$  and  $l$ , the maximal entanglement  $\Phi_d^{AR_A}$  must be obtained from  $|\Psi_k^p\rangle^{A'R_A B_1 B_E}$ . This transformation is equivalent to state merging from Bob to Alice. Consequently, any successful protocol is described as a combination of Markovianization of  $|\Psi\rangle$  and the subsequent state merging of  $|\Psi_k^p\rangle$  from Bob to Alice.

## VI. PROOF OF THEOREM 5

Let us return to the asymptotic scenario and prove Theorem 5. We consider protocols that transforms a state  $|\Psi^{\otimes n}\rangle^{\bar{A}\bar{B}\bar{R}_A\bar{R}_B} |\Phi_{K_n}\rangle^{A_0 B_0}$  into  $|\Phi_d^{\otimes n}\rangle^{\bar{A}\bar{R}_A} |\Phi_d^{\otimes n}\rangle^{\bar{B}\bar{R}_B}$   $|\Phi_{L_n}\rangle^{A_0 B_0}$ , as depicted in the right side of Figure 6. Conditions obtained in Section V directly apply by the following correspondence:

$$\begin{aligned} A, B, R_A, R_B &\rightarrow \bar{A}, \bar{B}, \bar{R}_A, \bar{R}_B, \quad U \rightarrow U^{\otimes n}, \quad \Phi_d \rightarrow \Phi_d^{\otimes n}, \\ \phi_{\text{res}} &\rightarrow \Phi_{K_n}, \quad \Psi \rightarrow \Psi^{\otimes n}, \quad \pi_d \rightarrow \pi_d^{\otimes n}. \end{aligned}$$

As presented in Section V, two-round protocols for this task is decomposed into two steps (see the left side of Figure 6). The first step is composed of Alice's measurement, forward classical communication and Bob's isometry. Markovianization by Alice's measurement satisfying the obliviousness condition is necessary and sufficient in order that Bob is able

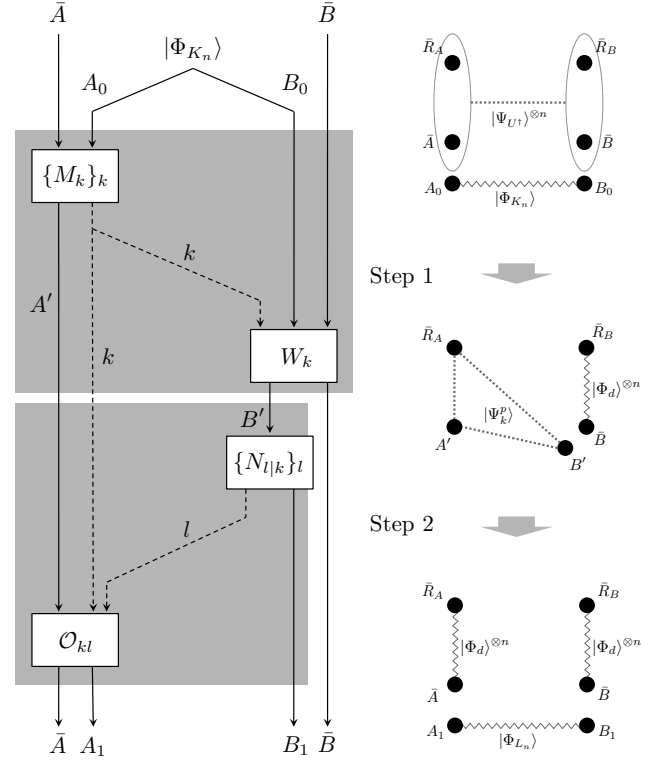


Fig. 6. The two-step protocol is depicted. Step 1: Markovianization by Alice's measurement and an isometric operation by Bob, aiming at obtaining  $|\Phi_d^{\otimes n}\rangle^{\bar{B}\bar{R}_B}$ , and Step 2: state merging from Bob to Alice, for obtaining  $|\Phi_d^{\otimes n}\rangle^{\bar{A}\bar{R}_A}$ . In the proof of achievability, we consider a case where  $A_1 = B_1 = \emptyset$  and  $\Psi_k^p = \tilde{\Psi}^p$ . In the proof of optimality,  $B'$  denotes a composite system  $B_1 B_E$ , and  $\{N_{l|k}\}_l$  is the projective measurement on  $B_E$  in the basis  $\{|l\rangle\}_l$ .

to obtain  $(\Phi_d^{BR_B})^{\otimes n}$ . The second step is composed of Bob's measurement, backward classical communication and Alice's local operation. To obtain  $(\Phi_d^{AR_A})^{\otimes n}$ , it is necessary and sufficient that the second step implements state merging of a particular tripartite state.

### A. Direct Part

We prove the direct part of Theorem 5. We assume  $L_n = 1$  in Definition 1, i.e., we consider a case where no entanglement is left after the protocol. The proof is by construction. Take arbitrary  $R > M(U^\dagger)$ , small  $\epsilon, r > 0$ , choose sufficiently large  $n$  and let  $K_n = 2^{n(R+r)}$ . Divide the resource state  $\Phi_{K_n}^{A_0 B_0}$  as

$$\Phi_{K_n}^{A_0 B_0} \rightarrow \Phi_{2^{nR}}^{A_0 B_0} \otimes \Phi_{2^{nr}}^{\tilde{A}_0 \tilde{B}_0}. \quad (29)$$

Consider a protocol consisting of the following steps.

- 1) **Alice's measurement:** By Definition 4, 8 and Theorem 10, there exists a random unitary operation  $\mathcal{V}_n : \tau \mapsto 2^{-nR} \sum_{j=1}^{2^{nR}} V_j \tau V_j^\dagger$  on  $\bar{A}$  such that  $\mathcal{V}_n(\Psi^{\otimes n})$  is  $\epsilon$ -recoverable from  $\bar{R}_A \bar{B} \bar{R}_B$ . Using  $V_j$  in  $\mathcal{V}_n$ , construct Alice's measurement  $\mathbb{M} = \{M_k^{AA_0 \rightarrow \bar{A}}\}_{k=1}^{2^{nR}}$  as

$$M_k^{\bar{A}A_0 \rightarrow \bar{A}} = \frac{1}{\sqrt{2^{nR}}} \sum_{j=1}^{2^{nR}} \exp\left(i \frac{2\pi j k}{2^{nR}}\right) \langle j|^{A_0} \otimes V_j^{\bar{A}}.$$



$\mathbb{M}$  is 0-oblivious and  $\epsilon$ -Markovianizing from  $\bar{R}_A \bar{B} \bar{R}_B$ . In addition, the reduced state of the post-measurement state on  $\bar{A} \bar{R}_A \bar{R}_B$  does not depend on  $k$ . Indeed, we have  $p_k = 2^{-nR}$  and

$$p_k^{-1} M_k(\tau^{\bar{A}} \otimes \Phi_{2^{nR}}^{A_0}) M_k^\dagger = \mathcal{V}_n(\tau)$$

for  $k = 1, \dots, 2^{nR}$ . Alice performs the measurement defined above.

- 2) **Forward classical communication:** Alice sends the measurement outcome  $k$  to Bob.
- 3) **Bob's isometry:** Due to Lemma 20, there exist pure states  $|\tilde{\Psi}^p\rangle^{\bar{A} \bar{R}_A \bar{B}}$ ,  $|\Psi'\rangle^{\bar{A} \bar{R}_A \bar{B} \bar{R}_B}$  and isometries  $\{W_k^{\bar{B} B_0 \rightarrow \bar{B} \bar{B}}\}_{k=1}^{2^{nR}}$  that satisfy

$$|\Psi'\rangle = p_k^{-1/2} (M_k \otimes W_k) |\Psi^{\otimes n}\rangle |\Phi_{K_n}\rangle$$

for any  $k$ , and satisfy

$$|\Psi'\rangle \approx |\tilde{\Psi}^p\rangle^{\bar{A} \bar{R}_A \bar{B}} |\Phi_d^{\otimes n}\rangle^{\bar{B} \bar{R}_B}, \quad (30)$$

$$(\tilde{\Psi}^p)^{\bar{R}_A} \approx \pi_d^{\bar{R}_A} \quad (31)$$

with a small error. Bob performs  $W_k$ .

- 4) **State merging:** Alice and Bob perform state merging of

$$|\tilde{\Psi}^p\rangle^{A' \bar{R}_A B'} := |\tilde{\Psi}^p\rangle^{\bar{A} \bar{R}_A \bar{B}} |\Phi_{2^{nr}}\rangle^{\bar{A}_0 \bar{B}_0}, \quad (32)$$

where  $A' = \bar{A} \bar{A}_0$  and  $B' = \bar{B} \bar{B}_0$ . Alice obtains a purification of  $(\tilde{\Psi}^p)^{\bar{R}_A}$  with a small error.

- 5) **Alice's isometry:** Alice performs an isometry and obtains  $|\Phi_d^{\otimes n}\rangle^{\bar{A} \bar{R}_A}$  within a small error.

The forward classical communication cost  $nC_f$  is simply equal to  $nR$  bits. As for Step 4), we consider a state merging in which no entanglement is obtained afterward. Thus the total entanglement cost is equal to the amount of entanglement that Alice and Bob have initially shared, i.e.,  $nE = n(R+r)$  ebits of (29). Applying Theorem 13 and the rank inequality in (28) for  $\tilde{\Psi}^p$ , the backward classical communication cost is bounded above by

$$\begin{aligned} nC_b &\leq \log \text{rank}[(\tilde{\Psi}^p)^{B'}] = \log \text{rank}[(\tilde{\Psi}^p)^{\bar{B}}] + nr \\ &\leq \log \dim[\mathcal{H}^{B_0}] + nr = n(R+r). \end{aligned}$$

In total, we have  $(E, C_f, C_b) = (R, R+r, R+r)$ .

The total error is evaluated by counting errors of (30), (31) and one induced by state merging (see Appendix E for the detail). Due to Lemma 21, the first two errors are bounded above by  $5\sqrt[4]{2\epsilon}$  and  $3\sqrt[4]{2\epsilon}$ , respectively. Theorem 13 implies that the merging error  $\epsilon_{\text{merg}}$  is bounded as

$$\epsilon_{\text{merg}} \leq 4 \cdot 2^{-nr/4}.$$

A simple calculation then yields an upper bound on the total error  $\epsilon_{\text{tot}}$ :

$$\epsilon_{\text{tot}} \leq 2\sqrt{3} \sqrt[8]{2\epsilon} + 5\sqrt[4]{2\epsilon} + 4 \cdot 2^{-nr/8}. \quad (33)$$

Since  $\epsilon, r > 0$  can be arbitrarily small, we conclude that a rate triplet  $(E, C_f, C_b) = (R, R, R)$  is achievable if  $R > M(U^\dagger)$ .

## B. Converse Part (Outline)

We prove the converse part of Theorem 5 by combining (16) and (19). Suppose a rate triplet  $(E, C_f, C_b)$  is achievable by a two-round protocol. By definition, for any  $\epsilon > 0$  and sufficiently large  $n$ , there exists a LOCC protocol  $\mathcal{M}_n$  that satisfies Condition (2). We then have

$$F(\rho(\mathcal{M}_n, U^\dagger), |\Phi_d^{\otimes n}\rangle^{\bar{A} \bar{R}_A} |\Phi_d^{\otimes n}\rangle^{\bar{B} \bar{R}_B} |\Phi_{L_n}\rangle^{A_1 B_1}) \geq 1 - \epsilon \quad (34)$$

for

$$\rho(\mathcal{M}_n, U^\dagger) := \mathcal{M}_n(|\Psi^{\otimes n}\rangle^{\bar{A} \bar{R}_A \bar{B} \bar{R}_B} |\Phi_{K_n}\rangle^{A_0 B_0}),$$

corresponding to (22). From Lemma 22, the map induced by Alice's measurement in  $\mathcal{M}_n$  is  $4\sqrt[4]{\epsilon}$ -oblivious and  $12\sqrt[4]{\epsilon}$ -Markovianizing from  $A' \bar{R}_A$ . Then the first two conditions in Definition 9 are satisfied by the following correspondence:

$$\begin{aligned} \bar{A}, \bar{B}, \bar{C}, A_0, G &\rightarrow \bar{A}, \bar{R}_A, (\bar{B} \bar{R}_B), A_0, B_0, \\ |\Psi\rangle^{ABC} &\rightarrow |\Psi\rangle^{A \bar{R}_A (B \bar{R}_B)}, \quad \phi_{\text{res}} \rightarrow \Phi_{K_n}, \\ \epsilon &\rightarrow 8\sqrt[4]{\epsilon}. \end{aligned} \quad (35)$$

In addition,  $\epsilon$  and  $n$  can be chosen for any small  $\delta > 0$  so that it satisfies

$$12\sqrt[4]{\epsilon} \cdot n \leq \delta, \quad (36)$$

due to Condition (7). Therefore, from (16), we have

$$\begin{aligned} H(\{p_k\}_k) &\geq \Delta S(A')_{av} \\ &\geq \Delta S(A')_{av} - \Delta S(B_0)_{av} \\ &\geq nM(U^\dagger) - n\tilde{\xi}(\delta). \end{aligned} \quad (37)$$

The optimality of the forward classical communication cost immediately follows from  $nC_f \geq H(\{p_k\}_{k \in \mathbb{K}})$ .

As for the backward classical communication cost, recall that Bob's measurement is decomposed into an isometry operation for obtaining  $(\Phi_d)^{\bar{B} \bar{R}_B}$  and a projective measurement on an ancillary system  $B_E$ . The latter forms state merging of  $\Psi_k^p$ , together with the backward classical communication and the subsequent Alice's local operation. Thus the backward classical communication cost is equal to the one required for performing state merging of  $\Psi_k^p$ . Due to (19) with the correspondence  $A \rightarrow A'$ ,  $B \rightarrow B_1 B_E$  and  $R \rightarrow \bar{R}_A$ , the cost is given by  $I(B_1 B_E : \bar{R}_A)_{\Psi_k^p}$ . Because of  $\Delta S(A')_{av} \gtrsim nM(U^\dagger)$  in (37), this cost turns out not to be smaller than  $nM(U^\dagger)$ .

The amount of entanglement obtained after state merging is bounded by (19) as  $\log L_n \lesssim -S(B_1 B_E | A')_{\Psi_k^p}$ , which implies the optimality of the total cost of entanglement when combined with  $\Delta S(A')_{av} - \Delta S(B_0)_{av} \gtrsim nM(U^\dagger)$ . See Appendix F for a detailed proof.

## VII. PROPERTIES OF THE COST

In this section, we investigate properties of the Markovianizing cost of unitaries. The results obtained here will be used in the next section for analyzing examples.

Consider a tripartite pure state  $|\Psi_U\rangle^{AR_A(BR_B)}$  defined by (6). The Petz recovery map  $\mathcal{R}_U : A \rightarrow A(BR_B)$  corresponding to  $|\Psi_U\rangle^{AR_A(BR_B)}$  is defined by

$$\mathcal{R}_U(\tau) = (\Psi_U^{A(BR_B)})^{\frac{1}{2}} (\Psi_U^A)^{-\frac{1}{2}} \tau (\Psi_U^A)^{-\frac{1}{2}} (\Psi_U^{A(BR_B)})^{\frac{1}{2}} \\ = U^{AB} (\text{Tr}_B[U^{\dagger AB} (\tau^A \otimes I^B) U^{AB}]) \otimes \Phi_d^{BR_B} U^{\dagger AB}$$

for  $\tau \in \mathcal{S}(\mathcal{H}^A)$  [16]. Define CPTP maps  $\mathcal{E}_U$  and  $\mathcal{E}_{U,\infty}$  on  $A$  by

$$\mathcal{E}_U := \text{Tr}_{BR_B} \circ \mathcal{R}_U, \quad \mathcal{E}_{U,\infty} := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathcal{E}_U^n, \quad (38)$$

and consider the states

$$\Psi_{U,\infty}^{AR_A BR_B} := \mathcal{E}_{U,\infty}^A(|\Psi_U\rangle\langle\Psi_U|^{AR_A BR_B}), \\ \Phi_{U,\infty}^{AR_A} := \mathcal{E}_{U,\infty}^A(|\Phi_d\rangle\langle\Phi_d|^{AR_A}). \quad (39)$$

As we prove below, the map  $\mathcal{E}_U$  is self-adjoint in the sense that  $\mathcal{E}_U = \mathcal{E}_U^*$ . Therefore, due to Theorem 9 in [12], the Markovianizing cost of  $U$  is given by

$$M(U) = M_{A|R_A}(\Psi_U^{AR_A(BR_B)}) = S(\Psi_{U,\infty}^{AR_A BR_B}),$$

which can be computed by a finite-step algorithm proposed in [12] (see Section III therein). Due to the unitary invariance of the von Neumann entropy, it immediately follows that

$$M(U) = S(\Phi_{U,\infty}^{AR_A}). \quad (40)$$

As a consequence, the Markovianizing costs of unitaries  $U_1$  and  $U_2$  are equal if they are local unitarily equivalent, that is, if there exist unitaries  $v, v'$  on  $\mathcal{H}^A$  and  $w, w'$  on  $\mathcal{H}^B$  such that  $U_1 = (v \otimes w) U_2 (v' \otimes w')$ .

Let us also analyze the Schmidt strength of unitaries. Consider Decomposition (8) of a bipartite unitary  $U$ . A CPTP map  $\mathcal{E}_U$  defined by (38) takes the form of

$$\mathcal{E}_U(\tau) = \sum_{ss'} c_s^2 c_{s'}^2 E_s^\dagger E_{s'} \tau E_{s'}^\dagger E_s = \sum_{ss'} \tilde{c}_{ss'}^2 \tilde{E}_{ss'}^\dagger \tau \tilde{E}_{ss'}, \quad (41)$$

where we introduced notations  $\tilde{c}_{ss'} := c_s c_{s'}$  and  $\tilde{E}_{ss'} := E_s^\dagger E_{s'}$ . It is straightforward to verify that  $\mathcal{E}_U$  is self-adjoint, that is, it satisfies

$$\mathcal{E}_U(\tau) = \mathcal{E}_U^*(\tau) := \sum_{ss'} \tilde{c}_{ss'}^2 \tilde{E}_{ss'}^\dagger \tau \tilde{E}_{ss'}.$$

Due to the orthonormality of  $\{E_s|\Phi_d\rangle^{AR_A}\}_s$  and  $\{F_s|\Phi_d\rangle^{BR_B}\}_s$ , which follows from (9), the eigen decomposition of  $\Psi_U^{AR_A}$  is given by

$$\Psi_U^{AR_A} = \sum_{s=0}^{d^2-1} c_s^2 E_s^A |\Phi_d\rangle\langle\Phi_d|^{AR_A} E_s^{\dagger A}. \quad (42)$$

Thus we have

$$K(U) = S(AR_A)_{\Psi_U}. \quad (43)$$

The following lemma provides a lower bound on the Markovianizing cost of unitaries.

*Lemma 24*  $M(U) \geq K(U)$  holds for any bipartite unitary  $U$ .

*Proof:* Define quantum operations  $e$  and  $e^*$  on  $\mathcal{S}(\mathcal{H}^A)$  by

$$e(\tau) = \sum_s c_s^2 E_s \tau E_s^\dagger, \quad e^*(\tau) = \sum_s c_s^2 E_s^\dagger \tau E_s.$$

From (42) and (43), the Schmidt strength of the unitary is given by

$$K(U) = S(e^A(|\Phi_d\rangle\langle\Phi_d|^{AR_A})). \quad (44)$$

It immediately follows from (41) that  $\mathcal{E}_U = e^* \circ e$ . We have

$$e(\tau) = \text{Tr}_B[U^{\dagger AB} (\tau^A \otimes I^B) U^{AB}] \\ e^*(\tau) = \text{Tr}_B[U^{AB} (\tau^A \otimes I^B) U^{\dagger AB}]$$

due to (8) and (9), which implies that  $e$  and  $e^*$  are unital, i.e.,  $e(I) = e^*(I) = I$ . Therefore, owing to the monotonicity of the von Neumann entropy under unital maps, we have

$$S((\mathcal{E}_U^n)^A(|\Phi_d\rangle\langle\Phi_d|^{AR_A})) \geq S(e^A(|\Phi_d\rangle\langle\Phi_d|^{AR_A})) \quad (45)$$

for any  $n \geq 1$ . Due to Definitions (38), (39) and the concavity of the von Neumann entropy, we obtain

$$S(\Phi_{U,\infty}^{AR_A}) \geq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N S((\mathcal{E}_U^n)^A(|\Phi_d\rangle\langle\Phi_d|^{AR_A})). \quad (46)$$

Expressions (40), (44), (45) and (46) yields  $M(U) \geq K(U)$ . ■

## VIII. EXAMPLES

In this section, we consider two classes of bipartite unitaries and compute their Markovianizing costs.

### A. Two-Qubit Unitaries

It is proved in [19] that all two-qubit unitaries are classified into the following categories:

- 1) Unitaries that can be written as a tensor product of local unitaries as  $U = u^A \otimes u^B$ . We do not consider this type of unitaries because of its triviality.
- 2) Unitaries that can be written in the form of

$$U = \cos\left(\frac{\theta}{2}\right) I^A \otimes I^B + i \sin\left(\frac{\theta}{2}\right) \sigma_z^A \otimes \sigma_z^B \quad (47)$$

up to local unitaries, where  $\theta \in (0, \pi/2]$ . Controlled-unitary gates are examples of such unitaries.

- 3) Unitaries that can be written in the form of

$$U = c_0 e^{i\theta_0} I^A \otimes I^B + c_1 e^{i\theta_1} \sigma_z^A \otimes \sigma_z^B \\ + c_2 e^{i\theta_2} \sigma_x^A \otimes \sigma_x^B + c_3 e^{i\theta_3} \sigma_y^A \otimes \sigma_y^B$$

up to local unitaries, where  $c_s, \theta_s \in \mathbb{R}$  ( $s = 0, 1, 2, 3$ ) are nonnegative real parameters satisfying  $\sum_{s=0}^3 c_s^2 = 1$ . All two-qubit unitaries that are not classified to the first two categories are of this category.

Let us consider unitaries of the form (47), which is local unitarily equivalent to the following controlled phase gate:

$$U_\theta = |0\rangle\langle 0|^A \otimes I^B + |1\rangle\langle 1|^A \otimes (e^{i\theta\sigma_z})^B.$$

We have

$$\begin{aligned} c_0 &= \cos(\theta/2), \quad c_1 = \sin(\theta/2), \\ \tilde{c}_{00}^2 + \tilde{c}_{11}^2 &= \frac{1}{2}(1 + \cos^2 \theta), \quad \tilde{c}_{01}^2 + \tilde{c}_{10}^2 = \frac{1}{2}\sin^2 \theta, \\ E_0 &= F_0 = I, \quad E_1 = \sigma_z, \quad F_1 = i\sigma_z, \\ \tilde{E}_{00} &= \tilde{E}_{11} = I, \quad \tilde{E}_{01} = \tilde{E}_{10} = \sigma_z. \end{aligned}$$

Thus a map corresponding to (41) is given by

$$\mathcal{E}(\tau) = \frac{1 + \cos^2 \theta}{2} \cdot \tau + \frac{1}{2} \sin^2 \theta \cdot \sigma_z \tau \sigma_z,$$

which leads to

$$\mathcal{E}_{U,\infty}(\tau) = \frac{1}{2}(\tau + \sigma_z \tau \sigma_z) = |0\rangle\langle 0| \tau |0\rangle\langle 0| + |1\rangle\langle 1| \tau |1\rangle\langle 1|$$

from (38). Hence we have

$$\Phi_{U,\infty}^{ARA} = \frac{1}{2}(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|)$$

corresponding to (39), which implies  $M(U) = 1$  due to (40). Consequently, we obtain the following theorem:

*Theorem 25* A rate triplet  $(E, C_f, C_b)$  is achievable by two-round protocols for implementing a two qubit controlled-unitary gate only if  $E, C_f, C_b \geq 1$ , if we additionally require in Definition 2 that Condition (7) holds.

The above theorem implies that, counterintuitively, at least 1 ebit of entanglement consumption per copy is necessary for implementing two-qubit controlled-unitary gate by two round protocols, regardless of how close the unitary is to the identity operation (i.e., regardless of how small  $\theta$  is). In [20], we prove that a certain class of two-qubit controlled-unitary gates can be implemented by a *four*-round protocol with the entanglement cost strictly smaller than 1 ebit per copy. Thereby we reveal a trade-off relation between the entanglement cost and the number of rounds for a LOCC task.

### B. Generalized Clifford Operators

The generalized Pauli operators  $\sigma_{pq}$  ( $p, q \in \{1, \dots, d\}$ ) on a  $d$ -dimensional Hilbert space is defined as

$$\begin{aligned} \sigma_{p0} &:= \sum_{t=1}^d |t-p\rangle\langle t|, \quad \sigma_{0q} := \sum_{t=1}^d e^{2\pi i q t/d} |t\rangle\langle t|, \\ \sigma_{pq} &:= \sigma_{p0} \sigma_{0q}, \end{aligned} \quad (48)$$

with a fixed basis  $\{|t\rangle\}_{t=1}^d$ . Here, subtraction is taken with mod  $d$ . A bipartite unitary  $U$  is called a generalized Clifford operator if, for any  $p, q, r$  and  $s$ , there exist  $p', q', r', s'$  and a phase  $\theta_{pqrs} \in \mathbb{R}$  such that

$$U(\sigma_{pq} \otimes \sigma_{rs})U^\dagger = e^{i\theta_{pqrs}} \sigma_{p'q'} \otimes \sigma_{r's'}.$$

The Markovianizing cost of generalized Clifford operators can be simply computed by the following theorem, a proof of which will be given in Appendix G.

*Theorem 26*  $M(U) = K(U)$  holds for any generalized Clifford operator  $U$ .

As a corollary of Theorem 5 and 26, the Schmidt strength  $K(U)$  is equal to the minimum cost of entanglement and classical communication for implementing generalized Clifford operator by two-round protocols under additional assumption (7). A stronger statement, represented by the following theorem, immediately follows from Theorem 6 and 26.

*Theorem 27* The following statements hold for any generalized Clifford operator  $U$  and  $r \geq 2$ .

- *Direct*: A rate triplet  $(E, C_f, C_b)$  is achievable by  $r$ -round protocols for implementing  $U$  if  $E, C_f, C_b \geq K(U)$ .
- *Converse*: A rate triplet  $(E, C_f, C_b)$  is achievable by  $r$ -round protocols for implementing  $U$  only if  $E, C_f, C_b \geq K(U)$ .

## IX. OPEN PROBLEM

We have derived a converse bound on the cost of entanglement and classical communication for implementing a bipartite unitary by two-round protocols. However, we do not know whether the converse bound remains to hold when we remove the additional requirement on the convergence speed of error, represented by Inequality (7). In this section, we investigate a relation between this open problem and another open problem regarding a property of approximate recoverability.

In the proof of the converse part, Condition (7) is exploited in the form of Inequality (36). This inequality is required to derive (37), in which an error term depends not on  $\epsilon$  but on  $\delta$ . The  $\delta$ -dependence of the error term originates from that in Inequality (16), and the latter arises due to the fact that Condition (15) is required to prove (14). In summary, we require Condition (7) to prove the converse part of Theorem 5 because we require Condition (15) to prove (14).

In [14], we proved that Condition (15) in Theorem 10 can be eliminated if a conjecture about approximate recoverability is true. The conjecture states that a tripartite quantum state  $\rho^{ABC}$  is approximately recoverable from  $\rho^{AB}$  by an operation from  $\mathcal{R} : B \rightarrow BC$  if it is approximately recoverable from  $\rho^{BC}$  by an operation  $\mathcal{R}' : B \rightarrow AB$ , up to a *dimension-independent* rescaling of error of recovery. A rigorous statement is as follows:

*Conjecture 28* (Conjecture 13 in [14]) There exists a nonnegative function  $g(\epsilon)$ , independent of the dimension of quantum systems and satisfies  $\lim_{\epsilon \rightarrow 0} g(\epsilon) = 0$ , such that the following statement holds for an arbitrary tripartite state  $\rho^{ABC}$  and  $\epsilon > 0$ : The state  $\rho^{ABC}$  is  $g(\epsilon)$ -recoverable from  $BC$  if it is  $\epsilon$ -recoverable from  $AB$ .

Condition (7) in Theorem 5 can be eliminated if the above conjecture is true. See also Appendix F-C.

## X. COMMUNICATION POWER OF A LOCC PROTOCOL

In this section, we analyze classical communication power of a LOCC protocol with an arbitrary preshared resource state. The results obtained here will be used in the next section to prove Theorem 6.

Consider the following scenario in which Alice aims to transmit  $nR$  bits of classical message to Bob by a bidirectional LOCC protocol that transforms a preshared quantum state  $\rho^{AB}$ .

- Alice and Bob initially share a bipartite quantum state  $\rho^{AB}$ .
- Alice is given an array of uniformly random classical bits  $\vec{X} = X_1 \cdots X_{nR}$ .
- Alice and Bob transforms  $\rho^{AB}$  by an LOCC protocol.
- Alice's operations during the protocol, as well as the message from Alice to Bob, may depend on  $\vec{X}$ .
- After the completion of the protocol, Bob performs a measurement on  $B$  to decode  $\vec{X}$ .

Let  $\vec{X}'$  be the result of Bob's decoding measurement. The decoding error is defined by

$$P_e := \Pr\{\vec{X} \neq \vec{X}'\}.$$

In the following, we prove that the length  $nR$  of classical message  $\vec{X}$  does not exceed the total number of classical bits transmitted from Alice to Bob during the protocol, if the decoding error is vanishingly small.

Without loss of generality, we assume that the protocol proceeds as follows. Here,  $\Gamma$  is a natural number, and  $K_\gamma$ ,  $L_\gamma$ ,  $\hat{K}_\gamma$ ,  $\hat{L}_\gamma$  are random variables which take values in finite sets  $\mathcal{K}_\gamma$ ,  $\mathcal{L}_\gamma$ ,  $\hat{\mathcal{K}}_\gamma$ ,  $\hat{\mathcal{L}}_\gamma$ , respectively.

- 1) Alice and Bob recursively apply the following operation from  $\gamma = 1$  to  $\gamma = \Gamma$ :
  - a) Alice performs a measurement  $\mathbb{M}_\gamma$  on her system and obtains an outcome  $K_\gamma$ .
  - b) Alice transmits a classical message  $\hat{K}_\gamma$  to Bob.
  - c) Bob performs a measurement  $\mathbb{N}_\gamma$  on his system and obtains an outcome  $L_\gamma$ .
  - d) Bob transmits a classical message  $\hat{L}_\gamma$  to Alice.

- 2) Alice performs a quantum operation on her system.

The total number of classical bits, transmitted from Alice to Bob during the protocol, is given by

$$C_{\text{tot}} := \sum_{\gamma=1}^{\Gamma} \log |\hat{\mathcal{K}}_\gamma|.$$

Let us introduce the following notations:

$$\begin{aligned} K^\gamma &:= (K_1, \dots, K_\gamma), & L^\gamma &:= (L_1, \dots, L_\gamma) \\ \hat{K}^\gamma &:= (\hat{K}_1, \dots, \hat{K}_\gamma), & \hat{L}^\gamma &:= (\hat{L}_1, \dots, \hat{L}_\gamma) \end{aligned}$$

In general, Alice and Bob's measurement in the protocol, as well as classical messages, may depend on the previous

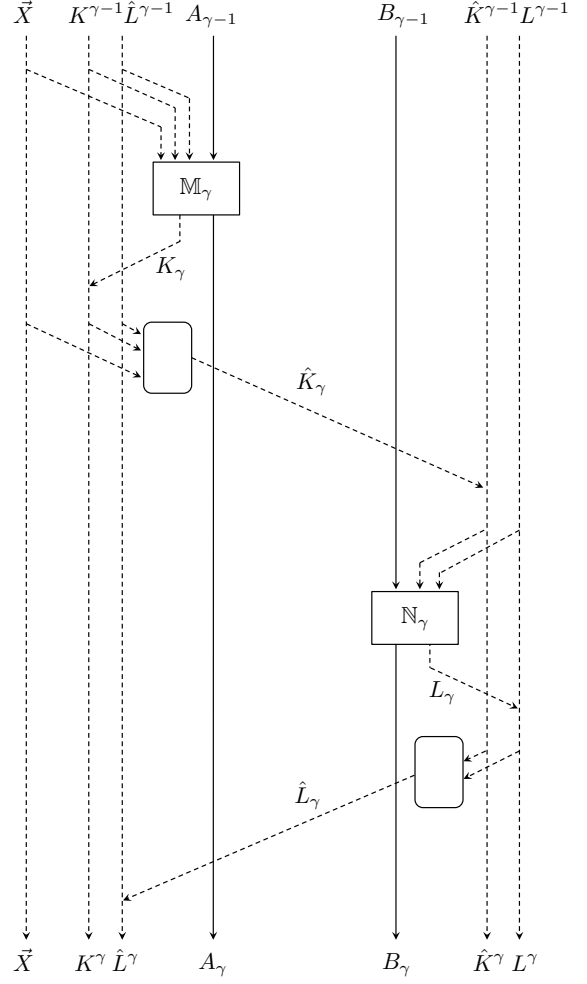


Fig. 7. A graphical representation of the  $\gamma$ -th step in an LOCC protocol is depicted. We denote system  $A$  and  $B$  after the  $\gamma$ -th step by  $A_\gamma$  and  $B_\gamma$  for  $\gamma = 1, \dots, \Gamma$ , respectively.

measurement outcomes and messages in the following way (Figure 7).

- $\mathbb{M}_\gamma$  depends on  $(\vec{X}, K^{\gamma-1}, \hat{L}^{\gamma-1})$ .
- $\mathbb{N}_\gamma$  depends on  $(L^{\gamma-1}, \hat{K}^{\gamma-1})$ .
- $\hat{K}_\gamma$  depends on  $(\vec{X}, K^\gamma, \hat{L}^{\gamma-1})$ .
- $\hat{L}_\gamma$  depends on  $(L^\gamma, \hat{K}^{\gamma-1})$ .

The following lemma states that the mutual information between  $\vec{X}$  and all that Bob has after the protocol is bounded above by the total amount of classical communication transmitted from Alice to Bob during the protocol. See Appendix H-A for a proof.

**Lemma 29** The following inequalities hold:

$$I(\vec{X} : B_\Gamma, L^\Gamma, \hat{K}^\Gamma) \leq C_{\text{tot}}, \quad (49)$$

$$nR \leq C_{\text{tot}} + h(P_e) + nRP_e. \quad (50)$$

Here,  $h(x)$  is the binary entropy defined by

$$h(x) := -x \log x - (1-x) \log (1-x),$$

and  $B_\Gamma$  denotes system  $B$  after the  $\Gamma$ -th step of the protocol.

*Remark.* An upper bound on the classical communication power of a two-way LOQC (local operations and *quantum* communication) protocol, which is similar to (50), has been proved in [21].

## XI. PROOF OF THEOREM 6

We prove Theorem 6 in this section, based on the idea that the cost of entanglement and classical communication for implementing a unitary is not smaller than powers of the unitary for generating entanglement and transmitting classical information ([10], [11], [22]).

Let us analyze power of a bipartite unitary for transmitting classical information. The following lemma states that the Schmidt strength is a lower bound on the classical communication power of a bipartite unitary. See Appendix H-B for a proof.

*Lemma 30* For any  $\epsilon \in (0, 1]$  and sufficiently large  $n$ , let  $\mathcal{U}_n$  be a quantum operation on  $A^n B^n$  that satisfy

$$F(\rho(\mathcal{U}_n), |\Psi_U\rangle^{\otimes n}) \geq 1 - \epsilon \quad (51)$$

for

$$\rho(\mathcal{U}_n) := \mathcal{U}_n(|\Phi_d^{ARA}\rangle^{\otimes n} |\Phi_d^{BRB}\rangle^{\otimes n}).$$

Then  $\mathcal{U}_n$  has a capacity to transmit  $n(K(U) - \epsilon)$  bits of classical information from Alice to Bob up to an error  $5\sqrt{\epsilon}$ , when assisted by shared entanglement.

Theorem 6 is then proved as follows.

*Proof of Theorem 6:* Suppose a rate triplet  $(E, C_f, C_b)$  is achievable. By definition, for any  $\epsilon > 0$  and sufficiently large  $n$ , there exist  $K_n$  and  $L_n$  that satisfy  $\log K_n - \log L_n = nE$ , and a LOCC protocol  $\mathcal{M}_n$  that satisfies (2), with the forward and backward classical communication cost  $nC_f$  and  $nC_b$ , respectively.

Define a quantum operation  $\hat{\mathcal{M}}_n$  on  $\bar{A}\bar{B}$  by

$$\hat{\mathcal{M}}_n : \tau \rightarrow \text{Tr}_{A_1 B_1} [\mathcal{M}_n(\tau^{\bar{A}\bar{B}} \otimes \Phi_{K_n}^{A_0 B_0})].$$

Due to Lemma 30,  $\hat{\mathcal{M}}_n$  has a capacity to transmit  $n(K(U) - \epsilon)$  bits of classical information from Alice to Bob up to an error  $5\sqrt{\epsilon}$ , when assisted by shared entanglement. By definition,  $\mathcal{M}_n$  has the same capacity. Applying Lemma 29 yields

$$n(K(U) - \epsilon) \leq nC_f + h(5\sqrt{\epsilon}) + n\epsilon(K(U) - \epsilon),$$

which leads to

$$(1 - \epsilon)(K(U) - \epsilon) \leq C_f + \frac{1}{n}h(5\sqrt{\epsilon}).$$

Since  $\epsilon > 0$  can be arbitrarily small, we obtain  $C_f \geq K(U)$ . Exchanging roles of Alice and Bob, we also have  $C_b \geq K(U)$ .

To prove  $E \geq K(U)$ , we assume for simplicity that  $K_n$  and  $L_n$  is bounded above as

$$\log K_n, \log L_n \leq n \log \kappa \quad (52)$$

with a constant  $\kappa > 0$ . We quantify entanglement of states between systems  $\bar{A}\bar{R}_A A_0$  and  $\bar{B}\bar{R}_B B_0$  (or between  $\bar{A}\bar{R}_A A_1$  and  $\bar{B}\bar{R}_B B_1$ ) by an entanglement measure that satisfy *asymptotic continuity* ([23], see Appendix A-D). We denote it by  $\hat{E}$ . Since  $\hat{E}$  is equal to the entanglement entropy for pure states, we have

$$\begin{aligned} \hat{E}(|\Psi_U\rangle^{\otimes n} |\Phi_{L_n}\rangle^{A_1 B_1}) &= nK(U) + \log L_n, \\ \hat{E}(|\Phi_d^{ARA}\rangle^{\otimes n} |\Phi_d^{BRB}\rangle^{\otimes n} |\Phi_{K_n}\rangle^{A_0 B_0}) &= \log K_n \end{aligned}$$

from (43). Due to asymptotic continuity, Condition (2) and (52) implies

$$\hat{E}(\rho(\mathcal{M}_n)) \geq nK(U) + \log L_n - n\delta(2\sqrt{\epsilon}) \log(d^4 \kappa^2), \quad (53)$$

where  $\delta(\epsilon)$  is an  $n$ -independent nonnegative function that satisfies  $\lim_{\epsilon \rightarrow 0} \delta(\epsilon) = 0$ . Equality (3) and the monotonicity of  $\hat{E}$  under LOCC operations yield

$$\hat{E}(\rho(\mathcal{M}_n)) \leq \log K_n. \quad (54)$$

Combining (53) and (54), we obtain

$$E = \frac{1}{n}(\log K_n - \log L_n) \geq K(U) - \delta(2\sqrt{\epsilon}) \log(d^4 \kappa^2),$$

which implies  $E \geq K(U)$  by taking the limit of  $\epsilon \rightarrow 0$ . ■

## XII. CONCLUSION

We have analyzed distributed quantum computation in terms of quantum Shannon theory for the first time. We have considered an asymptotic scenario for entanglement-assisted LOCC implementations of bipartite unitaries. For protocols consisting of two-round LOCC, we have derived the achievable rate region for the costs of entanglement and classical communication under an additional requirement on the convergence speed of error. We have also derived a general lower bound on the minimum cost of resources. The results can be straightforwardly generalized for cases where  $\dim \mathcal{H}^A \neq \dim \mathcal{H}^B$ . The problem formulated in this paper can be regarded as a quantum analog of ‘interactive coding for lossless computing’ in classical information theory [24].

## ACKNOWLEDGMENTS

This work is supported by the Project for Developing Innovation Systems of MEXT, Japan and JSPS KAKENHI (Grant No. 23540463, No. 23240001, No. 26330006, and No. 15H01677). We also gratefully acknowledge to the ELC project (Grant-in-Aid for Scientific Research on Innovative Areas MEXT KAKENHI (Grant No. 24106009)) for encouraging the research presented in this paper.

Some parts of the contents of this paper (Theorem 6, Theorem 27, Section XI, Appendix H-B and a part of Appendix G-B) were contained in our paper [25], which has

been submitted to IEEE Transactions on Information Theory and withdrawn afterward. The authors thank the reviewers of that paper for valuable comments, which has been useful in preparing this manuscript.

In [26] and the previous version of this manuscript, we failed to prove the converse part. The main weakness in the previous approach was that we exploit Markovianization in the version of [12], [13], rather than the one formulated in terms of approximate recoverability [14]. The authors thank the referees of ISIT 2015 for pointing out the relevance of approximate recoverability to the problem addressed in this paper.

## REFERENCES

- [1] J. Eisert, K. Jacobs, P. Papadopoulos, and M. Plenio, "Optimal local implementation of nonlocal quantum gates," *Phys. Rev. A*, vol. 62, p. 052317, 2000.
- [2] J. I. Cirac, W. Dur, B. Kraus, and M. Lewenstein, "Entangling operations and their implementation using a small amount of entanglement," *Phys. Rev. Lett.*, vol. 86, p. 544, 2001.
- [3] B. Groisman and B. Reznik, "Implementing nonlocal gates with non-maximally entangled states," *Phys. Rev. A*, vol. 71, p. 032322, 2005.
- [4] L. Chen and Y.-X. Chen, "Probabilistic implementation of a nonlocal operation using a nonmaximally entangled state," *Phys. Rev. A*, vol. 71, p. 054302, 2005.
- [5] M.-Y. Ye, Y.-S. Zhang, and G.-C. Guo, "Efficient implementation of controlled rotations by using entanglement," *Phys. Rev. A*, vol. 73, p. 032337, 2006.
- [6] D. W. Berry, "Implementation of multipartite unitary operations with limited resources," *Phys. Rev. A*, vol. 75, p. 032349, 2007.
- [7] N. B. Zhao and A. M. Wang, "Local implementation of nonlocal operations with block forms," *Phys. Rev. A*, vol. 78, p. 014305, 2008.
- [8] L. Yu, R. B. Griffiths, and S. M. Cohen, "Efficient implementation of bipartite nonlocal unitary gates using prior entanglement and classical communication," *Phys. Rev. A*, vol. 81, p. 062315, 2010.
- [9] S. M. Cohen, "Optimizing local protocols for implementing bipartite nonlocal unitary gates using prior entanglement and classical communication," *Phys. Rev. A*, vol. 81, p. 062316, 2010.
- [10] A. Soeda, P. Turner, and M. Murao, "Entanglement cost of implementing controlled-unitary operations," *Phys. Rev. Lett.*, vol. 107, p. 180501, 2011.
- [11] D. Stahlke and R. Griffiths, "Entanglement requirements for implementing bipartite unitary operations," *Phys. Rev. A*, vol. 84, p. 032316, 2011.
- [12] E. Wakakuwa, A. Soeda, and M. Murao, "Markovianizing cost of tripartite quantum states," e-print arXiv:1504.05805v3.
- [13] —, "Markovianizing cost of tripartite quantum states," in *Proc. IEEE International Symposium on Information Theory (ISIT 2015)*, Hong Kong, China, June 2015, pp. 700–704.
- [14] —, "The cost of randomness for converting a tripartite quantum state to be approximately recoverable," e-print arXiv:1512.06920v2, 2015.
- [15] B. Schumacher, "Quantum coding," *Phys. Rev. A*, vol. 51, p. 2738, 1995.
- [16] P. Hayden, R. Jozsa, D. Petz, and A. Winter, "Structure of states which satisfy strong subadditivity of quantum entropy with equality," *Comm. Math. Phys.*, vol. 246, pp. 359–374, 2004.
- [17] M. Horodecki, J. Oppenheim, and A. Winter, "Partial quantum information," *Nature*, vol. 436, pp. 673–676, 2005.
- [18] —, "Quantum state merging and negative information," *Comm. Math. Phys.*, vol. 269, pp. 107–136, 2007.
- [19] M. Nielsen, C. Dawson, J. Dodd, A. Gilchrist, D. Mortimer, T. Osborne, M. Bremner, A. Harrow, and A. Hines, "Quantum dynamics as a physical resource," *Phys. Rev. A*, vol. 67, p. 052301, 2003.
- [20] E. Wakakuwa, A. Soeda, and M. Murao, in preparation.
- [21] A. Nayak and J. Salzman, "On communication over an entanglement-assisted quantum channel," *Proc. of the 34th Ann. ACM Symp. on Theo. of Comp.*, pp. 698–704, 2002.
- [22] A. Montanaro and A. Winter, "A lower bound on entanglement-assisted quantum communication complexity," *ICALP'07 Proceedings of the 34th international conference on Automata, Languages and Programming*, pp. 122–133, 2007.
- [23] M. B. Plenio and S. Virmani, "An introduction to entanglement measures," *Quant. Inf. Comput.*, vol. 7, pp. 1–51, 2007.
- [24] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
- [25] E. Wakakuwa and M. Murao, "Asymptotic compressibility of entanglement and classical communication in distributed quantum computation," e-print arXiv:1310.3991.
- [26] E. Wakakuwa, A. Soeda, and M. Murao, "A coding theorem for distributed quantum computation," in *Proc. IEEE International Symposium on Information Theory (ISIT 2015)*, Hong Kong, China, June 2015, pp. 705–709.
- [27] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- [28] M. Hayashi, *Quantum Information: An Introduction*. Springer, 2006.
- [29] M. Wilde, *Quantum Information Theory*. Cambridge University Press, 2013.
- [30] A. Uhlmann, "The 'transition probability' in the state space of a \*-algebra," *Pep. Math. Phys.*, vol. 9, pp. 273–279, 1976.
- [31] M. Fannes, "A continuity property of the relative entropy density for spin lattice systems," *Comm. Math. Phys.*, vol. 31, pp. 291–294, 1973.
- [32] R. Alicki and M. Fannes, "Continuity of quantum conditional information," *J. Phys. A: Math. Gen.*, vol. 37.5, pp. L55–L57, 2004.
- [33] W. K. Wootters, "Entanglement of formation of an arbitrary state of two qubits," *Phys. Rev. Lett.*, vol. 80, p. 10, 1998.
- [34] M. A. Nielsen, "Continuity bounds for entanglement," *Phys. Rev. A*, vol. 61, p. 064301, 2000.
- [35] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, "Quantifying entanglement," *Phys. Rev. Lett.*, vol. 78, p. 2275, 1997.
- [36] V. Vedral and M. B. Plenio, "Entanglement measures and purification procedures," *Phys. Rev. A*, vol. 57, p. 1619, 1998.
- [37] M. J. Donald and M. Horodecki, "Continuity of relative entropy of entanglement," *Phys. Lett. A*, vol. 264, pp. 257–260, 1999.
- [38] M. Christandl and A. Winter, "'squashed entanglement': An additive entanglement measure," *Journal of Mathematical Physics*, vol. 45.3, pp. 829–840, 2004.
- [39] B. Groisman, S. Popescu, and A. Winter, "Quantum, classical, and total amount of correlations in a quantum state," *Phys. Rev. A*, vol. 72, p. 032317, 2005.
- [40] R. Ahlswede, "A method of coding and an application to arbitrarily varying channels," *J. Comb., Info. and Syst. Sciences*, vol. 5, pp. 10–35, 1980.
- [41] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, "Information causality as a physical principle," *Nature*, vol. 461, pp. 1101–1104, 2009.
- [42] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2006.
- [43] D. W. Berry, "Lower bounds for communication capacities of two-qudit unitary operations," *Phys. Rev. A*, vol. 76, p. 062302, 2007.
- [44] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channel," *Phys. Rev. A*, vol. 56, p. 131, 1997.
- [45] A. S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Trans. Inf. Theory*, vol. 44, pp. 269–273, 1998.

## APPENDIX A MATHEMATICAL PRELIMINARIES

In this appendix, we summarize technical tools that will be used in the following appendices. For the references, see e.g. [27]–[29]. See also Appendix A in [12] for basic properties of quantum entropies which are not presented here.

### A. Fidelity, Trace Distance and Uhlmann's Theorem

The trace distance between two quantum states  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  is defined by

$$\|\rho - \sigma\|_1 = \text{Tr} \left[ \sqrt{(\rho - \sigma)^2} \right].$$

It satisfies

$$0 \leq \|\rho - \sigma\|_1 \leq 2$$

and

$$\|\rho - \sigma\|_1 = 2 \max_{\Lambda} \text{Tr}[\Lambda(\rho - \sigma)], \quad (55)$$

where the maximization is taken over all linear operators  $\Lambda$  on  $\mathcal{H}$  that satisfy  $0 \leq \Lambda \leq I$ .

For  $\rho, \sigma, \tau \in \mathcal{S}(\mathcal{H})$ , we have

$$\|\rho - \tau\|_1 \leq \|\rho - \sigma\|_1 + \|\sigma - \tau\|_1, \quad (56)$$

which is called the *triangle inequality*. For two ensembles  $\{p_i, \rho_i\}$  and  $\{p_i, \sigma_i\}$ , we have

$$\sum_i p_i \|\rho_i - \sigma_i\|_1 \geq \left\| \sum_i p_i \rho_i - \sum_i p_i \sigma_i \right\|_1. \quad (57)$$

The trace distance takes a simple form under tensor product, i.e., for any  $\rho, \sigma \in \mathcal{S}(\mathcal{H}^A)$  and  $\tau \in \mathcal{S}(\mathcal{H}^B)$ , we have

$$\|\rho^A \otimes \tau^B - \sigma^A \otimes \tau^B\|_1 = \|\rho^A - \sigma^A\|_1. \quad (58)$$

The fidelity between two quantum states  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  is defined by

$$F(\rho, \sigma) := \left( \text{Tr} \left[ \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right] \right)^2,$$

and satisfies

$$0 \leq F(\rho, \sigma) \leq 1.$$

The fidelity takes a simple form for pure states as

$$F(|\psi\rangle, |\phi\rangle) = |\langle\psi|\phi\rangle|^2 \quad (59)$$

and

$$F(\rho, |\phi\rangle) = \langle\phi|\rho|\phi\rangle, \quad (60)$$

the latter of which yields

$$\sum_k p_k F(\rho_k, |\phi\rangle) = F\left(\sum_k p_k \rho_k, |\phi\rangle\right) \quad (61)$$

for any ensemble  $\{p_k, \rho_k\}_k$ . For

Let  $|\psi_\rho\rangle, |\psi_\sigma\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$  be arbitrary purifications of  $\rho, \sigma \in \mathcal{S}(\mathcal{H}^A)$ , respectively. Due to Uhlmann's theorem [30], we have

$$F(\rho, \sigma) = \max_W |\langle\psi_\rho|(I^A \otimes W^B)|\psi_\sigma\rangle|^2 \quad (62)$$

$$= \max_{\psi'_\sigma} |\langle\psi_\rho|\psi'_\sigma\rangle|^2. \quad (63)$$

Here, the maximization in the first line is taken over all unitaries  $W$  acting on  $\mathcal{H}^B$ , and that in the second line over all purifications  $|\psi'_\sigma\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$  of  $\sigma$ . It immediately follows that, for an arbitrary pure states  $|\Psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$  and  $|\phi\rangle \in \mathcal{H}^A$ , we have

$$F(\Psi^A, |\phi\rangle) = \max_{|\varphi\rangle \in \mathcal{H}^B} F(|\Psi\rangle^{AB}, |\phi\rangle^A |\varphi\rangle^B), \quad (64)$$

where the maximization is taken over all pure states on system  $B$ .

The trace distance and the fidelity are monotonic under quantum operations, i.e., it satisfies

$$\|\rho - \sigma\|_1 \geq \|\mathcal{E}(\rho) - \mathcal{E}(\sigma)\|_1, \quad (65)$$

$$F(\rho, \sigma) \leq F(\mathcal{E}(\rho), \mathcal{E}(\sigma))$$

for any  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  and any linear CPTP map  $\mathcal{E} : \mathcal{S}(\mathcal{H}) \rightarrow \mathcal{S}(\mathcal{H}')$ . In particular, the two functions are monotonic under taking the partial trace, that is, for any  $\rho, \sigma \in \mathcal{S}(\mathcal{H}^A \otimes \mathcal{H}^B)$  we have

$$\|\rho^{AB} - \sigma^{AB}\|_1 \geq \|\rho^A - \sigma^A\|_1, \quad (66)$$

$$F(\rho^{AB}, \sigma^{AB}) \leq F(\rho^A, \sigma^A). \quad (67)$$

The two functions are invariant under unitary operations, namely, for any unitary  $U$  acting on  $\mathcal{H}$  we have

$$\|\rho - \sigma\|_1 = \|U\rho U^\dagger - U\sigma U^\dagger\|_1, \quad (68)$$

$$F(\rho, \sigma) = F(U\rho U^\dagger, U\sigma U^\dagger). \quad (69)$$

The trace distance and the fidelity satisfy the following relation in general:

$$1 - \sqrt{F(\rho, \sigma)} \leq \frac{1}{2} \|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho, \sigma)}. \quad (70)$$

Therefore, if  $F(\rho, \sigma) \geq 1 - \epsilon$  then  $\|\rho - \sigma\|_1 \leq 2\sqrt{\epsilon}$ . Conversely, if  $\|\rho - \sigma\|_1 \leq \epsilon$  then  $F(\rho, \sigma) \geq 1 - \epsilon$ .

Let us introduce two lemmas that will be used in the following Appendices.

*Lemma 31* For any two bipartite pure states  $|\psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$  and  $|\phi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^{B'}$  that satisfy

$$\|\psi^A - \phi^A\|_1 \leq \epsilon, \quad (71)$$

the following statements hold:

- 1) There exists a linear CPTP map  $\mathcal{T} : B \rightarrow B'$  that satisfies

$$\|\mathcal{T}(|\psi\rangle\langle\psi|) - |\phi\rangle\langle\phi|\|_1 \leq 2\sqrt{\epsilon}. \quad (72)$$

- 2) If  $\dim \mathcal{H}^B \leq \dim \mathcal{H}^{B'}$ , then there exists an isometry  $\tilde{W} : \mathcal{H}^B \rightarrow \mathcal{H}^{B'}$  that satisfies

$$\|\tilde{W}|\psi\rangle\langle\psi|\tilde{W}^\dagger - |\phi\rangle\langle\phi|\|_1 \leq 2\sqrt{\epsilon}. \quad (73)$$

*Proof:* To prove 2), let  $|\tilde{\psi}\rangle \in \mathcal{H}^A \otimes \mathcal{H}^{B'}$  be a purification of  $\psi^A$ . Since all purifications are equivalent up to a local isometry, there exists an isometry  $W_1 : \mathcal{H}^B \rightarrow \mathcal{H}^{B'}$  that satisfies  $W_1|\psi\rangle = |\tilde{\psi}\rangle$ . From (71) and (70), the states satisfy

$$F(\tilde{\psi}^A, \phi^A) \geq 1 - \epsilon.$$

Due to (59) and (62), there exists a unitary  $W_2$  acting on  $\mathcal{H}^{B'}$  such that

$$F(W_2|\tilde{\psi}\rangle, |\phi\rangle) \geq 1 - \epsilon.$$

Using (70) once again, we obtain

$$\begin{aligned} & \|W_2 W_1 |\psi\rangle\langle\psi| W_1^\dagger W_2^\dagger - |\phi\rangle\langle\phi|\|_1 \\ &= \|W_2 |\tilde{\psi}\rangle\langle\tilde{\psi}| W_2^\dagger - |\phi\rangle\langle\phi|\|_1 \leq 2\sqrt{\epsilon}, \end{aligned}$$

which implies (73) by  $\tilde{W} := W_2 W_1$ .

To prove 1), let  $B''$  be an ancillary system such that

$$\dim \mathcal{H}^B \leq \dim \mathcal{H}^{B'} \times \dim \mathcal{H}^{B''}.$$

Due to 2), for any  $|\varphi\rangle \in \mathcal{H}^{B''}$  there exists an isometry  $\tilde{W} : \mathcal{H}^B \rightarrow \mathcal{H}^{B'} \otimes \mathcal{H}^{B''}$  that satisfies

$$\left\| \tilde{W} |\psi\rangle\langle\psi| \tilde{W}^\dagger - |\phi\rangle\langle\phi|^{AB'} \otimes |\varphi\rangle\langle\varphi|^{B''} \right\|_1 \leq 2\sqrt{\epsilon}. \quad (74)$$

Define a linear CPTP map  $\mathcal{T} : B \rightarrow B'$  by

$$\mathcal{T} : \tau \rightarrow \text{Tr}_{B''} [\tilde{W} \tau \tilde{W}^\dagger].$$

From (74) and (66), we obtain (72).  $\blacksquare$

### B. Gentle measurement lemma

The gentle measurement lemma (Lemma 9.4.1 in [29]) states that for any  $\rho \in \mathcal{S}(\mathcal{H})$ ,  $X \in \mathcal{L}(\mathcal{H})$  and  $\epsilon \geq 0$  such that  $0 \leq X \leq I$  and  $\text{Tr}[\rho X] \geq 1 - \epsilon$ , we have

$$\left\| \rho - \frac{\sqrt{X} \rho \sqrt{X}}{\text{Tr}[\rho X]} \right\|_1 \leq 2\sqrt{\epsilon}.$$

Let us introduce extensions of the gentle measurement lemma. Although similar lemmas have been used in the literature, we provide rigorous proofs for completeness.

*Lemma 32* For any  $\rho \in \mathcal{S}(\mathcal{H})$ ,  $X, Y \in \mathcal{L}(\mathcal{H})$  and  $\epsilon \in [0, 1]$  such that

$$0 \leq X \leq I, \quad 0 \leq Y \leq I$$

and

$$\text{Tr}[\rho X] \geq 1 - \epsilon, \quad \text{Tr}[\rho Y] \geq 1 - \epsilon, \quad (75)$$

define

$$\begin{aligned} D_{XY} &:= \text{Tr}[\sqrt{Y} \sqrt{X} \rho \sqrt{X} \sqrt{Y}], \\ \rho_{XY} &:= \frac{\sqrt{Y} \sqrt{X} \rho \sqrt{X} \sqrt{Y}}{D_{XY}}. \end{aligned}$$

Then we have

$$D_{XY} \geq 1 - 2\sqrt{\epsilon}, \quad \|\rho - \rho_{XY}\|_1 \leq 5\sqrt{\epsilon}.$$

*Proof:* Define

$$\rho_X := \frac{\sqrt{X} \rho \sqrt{X}}{\text{Tr}[\sqrt{X} \rho \sqrt{X}]}.$$

Due to the gentle measurement lemma, Condition (75) implies

$$\|\rho - \rho_X\|_1 \leq 2\sqrt{\epsilon}.$$

Consequently, we have

$$\begin{aligned} D_{XY} &= \text{Tr}[\rho_X Y] \\ &\geq \text{Tr}[\rho Y] - \frac{1}{2} \|\rho - \rho_X\|_1 \\ &\geq 1 - \epsilon - \sqrt{\epsilon} \geq 1 - 2\sqrt{\epsilon}, \end{aligned}$$

where the second line follows from (55), which leads to

$$\|\rho_X - \rho_{XY}\|_1 \leq 2\sqrt{2\sqrt{\epsilon}}$$

by (75). Thus we obtain

$$\begin{aligned} \|\rho - \rho_{XY}\|_1 &\leq \|\rho - \rho_X\|_1 + \|\rho_X - \rho_{XY}\|_1 \\ &\leq 2\sqrt{\epsilon} + 2\sqrt{2\sqrt{\epsilon}} \leq 5\sqrt[4]{\epsilon}. \end{aligned}$$

$\blacksquare$

*Lemma 33* Suppose  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  satisfy  $\|\rho - \sigma\|_1 \leq \epsilon$ . Let  $\{\lambda_i^\downarrow\}_{i=1}^d$  be the eigenvalues of  $\sigma$  sorted in decreasing order, where  $d := \dim \mathcal{H}$ , and let

$$\sigma = \sum_{i=1}^d \lambda_i^\downarrow |i\rangle\langle i|$$

be the eigen decomposition of  $\sigma$ . Define a projection operator

$$\tilde{\Pi} := \sum_{i=1}^{\text{rank}[\rho]} |i\rangle\langle i|.$$

Then we have

$$\text{Tr}[\tilde{\Pi} \sigma] \geq 1 - \epsilon.$$

*Proof:* Let  $\Pi_\rho$  be the projection onto  $\text{supp}[\rho]$ , and  $\Pi_\rho^\perp$  be that onto its orthogonal complement (i.e.,  $\Pi_\rho^\perp = I - \Pi_\rho$ ). Then we have

$$\begin{aligned} \epsilon &\geq \|\rho - \sigma\|_1 \geq \|\rho - \Pi_\rho \sigma \Pi_\rho - \Pi_\rho^\perp \sigma \Pi_\rho^\perp\|_1 \\ &= \|\rho - \Pi_\rho \sigma \Pi_\rho\|_1 + \text{Tr}[\Pi_\rho^\perp \sigma] \geq 1 - \text{Tr}[\Pi_\rho \sigma], \end{aligned} \quad (76)$$

where the second inequality follows from the monotonicity of the trace distance under a linear CPTP map defined by

$$\tau \rightarrow \Pi_\rho \tau \Pi_\rho + \Pi_\rho^\perp \tau \Pi_\rho^\perp.$$

We also have

$$\begin{aligned} \text{Tr}[\tilde{\Pi} \sigma] &= \sum_{i=1}^{\text{rank}[\rho]} \lambda_i^\downarrow \\ &= \sum_{i=1}^{\text{rank}[\rho]} \lambda_i^\downarrow \langle i | \Pi_\rho | i \rangle + \sum_{i=1}^{\text{rank}[\rho]} \lambda_i^\downarrow (1 - \langle i | \Pi_\rho | i \rangle) \\ &\geq \sum_{i=1}^{\text{rank}[\rho]} \lambda_i^\downarrow \langle i | \Pi_\rho | i \rangle + \lambda_{\text{rank}[\rho]}^\downarrow \sum_{i=1}^{\text{rank}[\rho]} (1 - \langle i | \Pi_\rho | i \rangle) \\ &= \sum_{i=1}^{\text{rank}[\rho]} \lambda_i^\downarrow \langle i | \Pi_\rho | i \rangle + \lambda_{\text{rank}[\rho]}^\downarrow \sum_{i=\text{rank}[\rho]+1}^d \langle i | \Pi_\rho | i \rangle \\ &\geq \sum_{i=1}^d \lambda_i^\downarrow \langle i | \Pi_\rho | i \rangle = \text{Tr}[\Pi_\rho \sigma], \end{aligned} \quad (77)$$

where the fourth line follows due to

$$\sum_{i=1}^d \langle i | \Pi_\rho | i \rangle = \text{Tr}[\Pi_\rho] = \text{rank}[\rho].$$

From (76) and (77), we obtain

$$\text{Tr}[\tilde{\Pi} \sigma] \geq \text{Tr}[\Pi_\rho \sigma] \geq 1 - \epsilon. \quad \blacksquare$$



### C. Continuity of Quantum Entropies

Define

$$\eta_0(x) := \begin{cases} -x \log x & (x \leq 1/e) \\ \frac{1}{e} & (x \geq 1/e) \end{cases}$$

and  $\eta(x) = x + \eta_0(x)$ , where  $e$  is the base of the natural logarithm. Define also

$$h(x) := -x \log x - (1-x) \log(1-x).$$

For two states  $\rho$  and  $\sigma$  in a  $d$ -dimensional quantum system ( $d < \infty$ ) such that  $\|\rho - \sigma\|_1 \leq \epsilon$ , we have

$$|S(\rho) - S(\sigma)| \leq \epsilon \log d + \eta_0(\epsilon), \quad (78)$$

which is called the *Fannes inequality* [31]. A simple calculation then yields

$$|S(\rho) - S(\sigma)| \leq \eta(\epsilon) \log d. \quad (79)$$

For two bipartite states  $\rho, \sigma \in \mathcal{S}(\mathcal{H}^A \otimes \mathcal{H}^B)$  such that  $\|\rho - \sigma\|_1 \leq \epsilon < 1$ , we have

$$|S(A|B)_\rho - S(A|B)_\sigma| \leq 4\epsilon \log d_A + 2h(\epsilon).$$

This inequality is called the *Alicki-Fannes inequality* [32], and leads to

$$|S(A|B)_\rho - S(A|B)_\sigma| \leq 4\eta(\epsilon) \log d_A. \quad (80)$$

Note that the upper bound in (80) does not depend on  $d_B$ . As a consequence, we have

$$\begin{aligned} & |I(A : B)_\rho - I(A : B)_\sigma| \\ & \leq |S(A)_\rho - S(A)_\sigma| + |S(A|B)_\rho - S(A|B)_\sigma| \\ & \leq 5\eta(\epsilon) \log d_A. \end{aligned} \quad (81)$$

The following lemma will be used for evaluating average errors.

*Lemma 34* Let  $c \in (0, \infty)$  be a constant,  $f : [0, c] \rightarrow \mathbb{R}$  be a monotonically nondecreasing function that satisfies  $f(c) < \infty$ , and  $\{p_k\}_{k \in \mathbb{K}}$  be a probability distribution on a countable set  $\mathbb{K}$ . Suppose  $\epsilon_k$  ( $k \in \mathbb{K}$ ) satisfies  $\epsilon_k \in [0, c]$ , and  $\sum_{k \in \mathbb{K}} p_k \epsilon_k \leq \epsilon$  for a given  $\epsilon \in (0, c^2]$ . Then we have

$$\sum_{k \in \mathbb{K}} p_k f(\epsilon_k) \leq f(\sqrt{\epsilon}) + f(c) \cdot \sqrt{\epsilon}. \quad (82)$$

*Proof:* Define  $\mathbb{K}(\lambda) := \{k \in \mathbb{K} \mid \epsilon_k \leq \lambda\}$ . For any  $t > 0$ , we have

$$\begin{aligned} \sum_{k \in \mathbb{K} \setminus \mathbb{K}(t\epsilon)} p_k &= \frac{1}{t\epsilon} \sum_{k \in \mathbb{K} \setminus \mathbb{K}(t\epsilon)} p_k t\epsilon \leq \frac{1}{t\epsilon} \sum_{k \in \mathbb{K} \setminus \mathbb{K}(t\epsilon)} p_k \epsilon_k \\ &\leq \frac{1}{t\epsilon} \sum_{k \in \mathbb{K}} p_k \epsilon_k \leq \frac{1}{t}, \end{aligned}$$

and consequently,

$$\begin{aligned} \sum_{k \in \mathbb{K}} p_k f(\epsilon_k) &= \sum_{k \in \mathbb{K}(t\epsilon)} p_k f(\epsilon_k) + \sum_{k \in \mathbb{K} \setminus \mathbb{K}(t\epsilon)} p_k f(\epsilon_k) \\ &\leq f(t\epsilon) + \frac{f(c)}{t}. \end{aligned}$$

Choosing  $t = 1/\sqrt{\epsilon}$ , we obtain (82). ■

### D. Entanglement Measures

A function  $\hat{E} : \mathcal{S}(\mathcal{H}^A \otimes \mathcal{H}^B) \rightarrow [0, \infty)$  is called an *entanglement measure* if it satisfies the following three properties [23]:

- 1) If  $\rho$  is a pure state on  $AB$ , then  $\hat{E}(\rho) = S(\rho^A)$ .
- 2) If  $\rho$  is a separable state on  $AB$ , then  $\hat{E}(\rho) = 0$ .
- 3)  $\hat{E}(\rho)$  does not increase on average under LOCC, i.e., if an ensemble  $\{p_i, \rho_i\}$  ( $\rho_i \in \mathcal{S}(\mathcal{H}^A \otimes \mathcal{H}^B)$ ) is obtained from  $\rho \in \mathcal{S}(\mathcal{H}^A \otimes \mathcal{H}^B)$  by an LOCC transformation between  $A$  and  $B$ , then  $\sum_i p_i \hat{E}(\rho_i) \leq \hat{E}(\rho)$ .

An entanglement measure  $\hat{E}$  is said to be *asymptotically continuous*, if there exists an  $n$ -independent nonnegative function  $\delta(\epsilon)$  that satisfies  $\lim_{\epsilon \rightarrow 0} \delta(\epsilon) = 0$ , and it holds that

$$\left| \hat{E}(\rho) - \hat{E}(\sigma) \right| \leq \delta(\epsilon) \log d_A d_B$$

for all  $\rho, \sigma \in \mathcal{S}(\mathcal{H}^A \otimes \mathcal{H}^B)$  satisfying  $\|\rho - \sigma\|_1 \leq \epsilon$ . Examples of asymptotically continuous entanglement measures are entanglement of formation [33], [34], the relative entropy of entanglement [35]–[37] and squashed entanglement [32], [38].

## APPENDIX B

### PROOFS OF INEQUALITY (4), LEMMA 11 AND INEQUALITIES (19)

#### A. Proof of Inequality (4)

We prove that Condition (2) implies (4). Due to the unitary invariance of the fidelity (69), we have

$$\begin{aligned} & F(\rho(\mathcal{M}_n, \psi), (U^{\otimes n} |\psi\rangle)^{\bar{A}\bar{B}} |\Phi_{L_n}\rangle^{A_0 B_0}) \\ &= F(U^{\dagger \otimes n} \rho(\mathcal{M}_n, \psi) U^{\otimes n}, |\psi\rangle^{\bar{A}\bar{B}} |\Phi_{L_n}\rangle^{A_0 B_0}) \end{aligned} \quad (83)$$

and

$$\begin{aligned} & F(\rho(\mathcal{M}_n), |\Psi_U\rangle^{\otimes n} |\Phi_{L_n}\rangle^{A_1 B_1}) \\ &= F(U^{\dagger \otimes n} \rho(\mathcal{M}_n) U^{\otimes n}, |\Phi_d^{AR_A}\rangle^{\otimes n} |\Phi_d^{BR_B}\rangle^{\otimes n} |\Phi_{L_n}\rangle^{A_1 B_1}). \end{aligned} \quad (84)$$

Since the “expected fidelity” is always greater than or equal to the “entanglement fidelity” (see e.g. Section 9.5 of [29]), we have

$$\begin{aligned} & \int_{\text{Haar}} p(d\psi) F(U^{\dagger \otimes n} \rho(\mathcal{M}_n, \psi) U^{\otimes n}, |\psi\rangle^{\bar{A}\bar{B}} |\Phi_{L_n}\rangle^{A_0 B_0}) \\ & \geq F(U^{\dagger \otimes n} \rho(\mathcal{M}_n) U^{\otimes n}, |\Phi_d^{AR_A}\rangle^{\otimes n} |\Phi_d^{BR_B}\rangle^{\otimes n} |\Phi_{L_n}\rangle^{A_1 B_1}). \end{aligned} \quad (85)$$

Inequalities (83), (84), (85) and Condition (2) implies (4). ■

#### B. Proof of Lemma 11

Let  $V : \bar{A}A_0 \rightarrow A'E_0$  be an isometry such that the Naimark extension of  $\{M_k\}_{k \in \mathbb{K}}$  is given by  $M_k = \langle k|^{E_0} V$ , and let

$$\Psi'^{A'E_0} := \sum_{k \in \mathbb{K}} |k\rangle\langle k|^{E_0} V ((\Psi^{\otimes n})^{\bar{A}} \otimes \varrho_n^{A_0}) V^\dagger |k\rangle\langle k|^{E_0}.$$

We have

$$\begin{aligned} H(\{p_k\}_{k \in \mathbb{K}}) &= S(E_0)_{\Psi'} = S(A'E_0)_{\Psi'} - S(A'|E_0)_{\Psi'} \\ &\geq S(\bar{A}A_0)_{\Psi^{\otimes n} \otimes \varrho_n} - \sum_{k \in \mathbb{K}} p_k S(A')_{\rho_k} = \Delta S(A')_{ave}, \end{aligned}$$

where the second line follows due to the von Neumann entropy nondecreasing under dephasing operations. Hence we obtain the first inequality in (16). The second inequality is due to  $\Delta S(G)_{av} \geq 0$ , which follows from the subadditivity of the von Neumann entropy.

As for the third inequality, we first prove that there exists a nondecreasing function  $\tilde{\eta}(\epsilon)$ , satisfying  $\lim_{\epsilon \rightarrow 0} \tilde{\eta}(\epsilon) = 0$ , such that

$$\begin{aligned} \Delta S(A')_{ave} - \Delta S(G)_{ave} \\ \geq I(\bar{B}\bar{C} : G)_{ave} - n\tilde{\eta}(\epsilon) \log(d_B d_C). \end{aligned}$$

Define

$$\epsilon_k := \left\| (\Psi^{\otimes n})^{\bar{B}\bar{C}} - \Psi_k^{\bar{B}\bar{C}} \right\|_1.$$

Using (78), we have

$$\begin{aligned} \Delta S(A')_k &= nS(A)_{\Psi} + S(A_0)_{\varrho_n} - S(A')_{\Psi_k} \\ &= S(\bar{B}\bar{C})_{\Psi^{\otimes n}} + S(A_0)_{\varrho_n} - S(\bar{B}\bar{C}G)_{\Psi_k} \\ &\geq S(\bar{B}\bar{C})_{\Psi_k} + S(A_0)_{\varrho_n} - S(\bar{B}\bar{C}G)_{\Psi_k} \\ &\quad - n\eta(\epsilon_k) \log(d_B d_C) \\ &= S(A_0)_{\varrho_n} - S(G|\bar{B}\bar{C})_{\Psi_k} - n\eta(\epsilon_k) \log(d_B d_C) \\ &= S(G)_{\varrho_n} - S(G)_{\Psi_k} + I(\bar{B}\bar{C} : G)_{\Psi_k} \\ &\quad - n\eta(\epsilon_k) \log(d_B d_C) \\ &= \Delta S(G)_k + I(\bar{B}\bar{C} : G)_{\Psi_k} - n\eta(\epsilon_k) \log(d_B d_C), \end{aligned}$$

where, in the fifth line, we used the fact that  $\varrho_n$  is a pure state on  $A_0G$ . Averaging over  $k$ , we obtain

$$\begin{aligned} \Delta S(A')_{ave} - \Delta S(G)_{ave} \\ \geq \sum_{k \in \mathbb{K}} p_k (I(\bar{B}\bar{C} : G)_{\Psi_k} - n\eta(\epsilon_k) \log(d_B d_C)). \end{aligned}$$

Applying Lemma 34 together with  $\sum_{k \in \mathbb{K}} p_k \epsilon_k \leq \epsilon$  and  $\epsilon_k \leq 2$  yields

$$\begin{aligned} \Delta S(A')_{ave} - \Delta S(G)_{ave} \\ \geq I(\bar{B}\bar{C} : G)_{ave} - n\tilde{\eta}(\epsilon) \log(d_B d_C), \end{aligned}$$

where we defined

$$\tilde{\eta}(\epsilon) := \eta(\sqrt{\epsilon}) + \eta(2) \cdot \sqrt{\epsilon}.$$

Second we prove that there exists a function  $\xi(\delta)$ , satisfying  $\lim_{\delta \rightarrow 0} \xi(\delta) = 0$ , such that we have

$$I(\bar{B}\bar{C} : G)_{ave} \geq M_{A|AB}^{R,m}(\Psi^{ABC}) - n\xi(\delta) \log(d_A d_B d_C). \quad (86)$$

This simply follows from the results in [14] (see Theorem 15 and Inequality (66) therein). Defining

$$\tilde{\xi}(\delta) := \xi(\delta) + \tilde{\eta}(\delta), \quad (87)$$

and noting  $\tilde{\eta}(\epsilon) \leq \tilde{\eta}(\delta)$ , we obtain the last inequality in (16). ■

### C. Proof of Inequalities (19)

The following theorem is essentially the same, but technically different from what is proved in [17]. We give a rigorous proof for completeness.

*Theorem 35* Let  $\mathcal{N}$  be state merging of  $\Psi$  with error  $\epsilon \in (0, 1/4]$ . Entanglement cost and classical communication cost of  $\mathcal{N}$  are bounded below as

$$\begin{aligned} \log K - \log L &\geq S(B|A)_{\Psi} - \eta'(2\sqrt{\epsilon}) \log(d_R L), \\ C &\geq I(B : R)_{\Psi} - 5\eta(2\sqrt{\epsilon}) \log d_R, \end{aligned}$$

where

$$\eta'(x) := \frac{5}{2}\eta(x) + \eta(2\sqrt{x}) + \eta(2)\sqrt{x}. \quad (88)$$

*Proof:* Without loss of generality, we assume that the protocol  $\mathcal{N}$  consists of (i) Bob's measurement described by  $\{N_l^{BB_0 \rightarrow B_1}\}_l$ , (ii) communication of  $l$  from Bob to Alice, and (iii) Alice's operation described by a CPTP map  $\mathcal{O}_l : AA_0 \rightarrow AA_B A_1$ . The final state is given by

$$\rho(\mathcal{N})^{AA_B RA_1 B_1} = \sum_l p_l \hat{\Psi}_l, \quad (89)$$

where

$$p_l := \|N_l|\Psi\rangle\langle\Phi_K|\|_1^2, \quad |\Psi_l\rangle := p_l^{-1/2} N_l|\Psi\rangle\langle\Phi_K|$$

and  $\hat{\Psi}_l := \mathcal{O}_l(|\Psi_l\rangle)$ . From (17) and (70), we have

$$\|\rho(\mathcal{N}) - \Psi^{AA_B R} \otimes \Phi_L^{A_1 B_1}\|_1 \leq 2\sqrt{\epsilon}. \quad (90)$$

Define

$$\epsilon_l := \left\| \hat{\Psi}_l - \Psi^{AA_B R} \otimes \Phi_L^{A_1 B_1} \right\|_1 \quad (91)$$

and

$$f_l := F(\hat{\Psi}_l, \Psi^{AA_B R} \otimes \Phi_L^{A_1 B_1})$$

for each  $l$ . Due to the convexity of the square function, Inequality (70), Equalities (61), (89) and Inequality (17), we have

$$\begin{aligned} \left( \sum_l p_l \epsilon_l \right)^2 &\leq \sum_l p_l \epsilon_l^2 \leq 4 \sum_l p_l (1 - f_l) \\ &= 4 - 4F(\rho(\mathcal{N}), |\Psi\rangle^{AA_B R} |\Phi_L\rangle^{A_1 B_1}) \leq 4\epsilon, \end{aligned}$$

which yields

$$\sum_l p_l \epsilon_l \leq 2\sqrt{\epsilon}.$$

Consider the following protocol, which is as a whole equivalent to the protocol described above.

- 1) Bob performs a CPTP map  $\mathcal{E}_1 : BB_0 \rightarrow B_1 C$  defined by  $\mathcal{E}_1(\tau) = \sum_l |l\rangle\langle l|^C \otimes N_l \tau N_l^\dagger$ . The state after this operation is  $\Psi' = \sum_l p_l |l\rangle\langle l|^C \otimes |\Psi_l\rangle\langle\Psi_l|^{AA_0 B_1 R}$ .
- 2) Bob transmits system  $C$  to Alice.
- 3) Alice performs a CPTP map  $\mathcal{E}_2 : CAA_0 \rightarrow AA_B A_1$  defined as  $\mathcal{E}_2(\tau) = \sum_l \mathcal{O}_l(|l\rangle\langle l|^C \tau^{CAA_0} |l\rangle\langle l|^C)$ . The state after the operation is  $\mathcal{E}_2(\Psi') = \rho(\mathcal{N})$ .

By the chain rule and the data processing inequality, we have

$$\begin{aligned}
& 2S(A)_\Psi + 2\log K \\
&= I(AA_0 : BB_0R)_{\Psi \otimes \Phi_K} \\
&\geq I(AA_0 : B_1CR)_{\Psi'} \\
&= I(AA_0 : C)_{\Psi'} + I(AA_0 : B_1R|C)_{\Psi'} \\
&\geq I(AA_0C : B_1R)_{\Psi'} - I(C : B_1R)_{\Psi'} \\
&\geq I(AA_BA_1 : B_1R)_{\rho(\mathcal{N})} - I(C : B_1R)_{\Psi'}. \quad (92)
\end{aligned}$$

Due to Inequality (90) and (81), we have

$$\begin{aligned}
& I(AA_BA_1 : B_1R)_{\rho(\mathcal{N})} \\
&\geq I(AA_BA_1 : B_1R)_{\Psi \otimes \Phi_L} - 5\eta(2\sqrt{\epsilon}) \log(d_R L) \\
&= I(AA_B : R)_\Psi + I(A_1 : B_1)_{\Phi_L} - 5\eta(2\sqrt{\epsilon}) \log(d_R L) \\
&= 2S(R)_\Psi + 2\log L - 5\eta(2\sqrt{\epsilon}) \log(d_R L). \quad (93)
\end{aligned}$$

From (90), (91), (81) and Lemma 34, we also have

$$\begin{aligned}
& I(C : B_1R)_{\Psi'} \\
&= S(B_1R)_{\Psi'} - S(B_1R|C)_{\Psi'} \\
&= S(B_1R)_{\rho(\mathcal{N})} - \sum_l p_l S(B_1R)_{\hat{\Psi}_l} \\
&= S(B_1R)_{\rho(\mathcal{N})} - S(B_1R)_{\Psi \otimes \Phi_L} \\
&\quad + \sum_l p_l \left( S(B_1R)_{\Psi \otimes \Phi_L} - S(B_1R)_{\hat{\Psi}_l} \right) \\
&\leq \sum_l p_l (\eta(\epsilon_l) + \eta(2\sqrt{\epsilon_l})) \log(d_R L) \\
&\leq 2 \sum_l p_l \eta(2\sqrt{\epsilon_l}) \log(d_R L) \\
&\leq 2 \left( \eta(2\sqrt{\epsilon}) + \eta(2) \cdot \sqrt{2\sqrt{\epsilon}} \right) \log(d_R L). \quad (94)
\end{aligned}$$

From (92), (93) and (94), we obtain

$$\begin{aligned}
& \log K - \log L \\
&\geq S(R)_\Psi - S(A)_\Psi - \eta'(2\sqrt{\epsilon}) \log(d_R L) \\
&= S(AB)_\Psi - S(A)_\Psi - \eta'(2\sqrt{\epsilon}) \log(d_R L) \\
&= S(B|A)_\Psi - \eta'(2\sqrt{\epsilon}) \log(d_R L)
\end{aligned}$$

for the entanglement cost. As for the classical communication cost, from (90) and (81), we have

$$\begin{aligned}
& 2S(R)_\Psi \\
&= I(AA_B : R)_\Psi \\
&\leq I(AA_B : R)_{\rho(\mathcal{N})} + 5\eta(2\sqrt{\epsilon}) \log d_R \\
&\leq I(AA_0C : R)_{\Psi'} + 5\eta(2\sqrt{\epsilon}) \log d_R \\
&= I(AA_0 : R)_{\Psi'} + I(C : R|AA_0)_{\Psi'} + 5\eta(2\sqrt{\epsilon}) \log d_R \\
&= I(AA_0 : R)_{\Psi \otimes \Phi_K} + I(C : AA_0R)_{\Psi'} \\
&\quad - I(C : AA_0)_{\Psi'} + 5\eta(2\sqrt{\epsilon}) \log d_R \\
&\leq I(A : R)_\Psi + S(C)_{\Psi'} + 5\eta(2\sqrt{\epsilon}) \log d_R \\
&= I(A : R)_\Psi + H(\{p_l\}_l) + 5\eta(2\sqrt{\epsilon}) \log d_R.
\end{aligned}$$

Here, the fifth line follows from the fact that Bob's measurement does not change the average reduced state of  $AA_0R$ .

Thus we obtain

$$\begin{aligned}
C &\geq H(\{p_l\}_l) \\
&\geq 2S(R)_\Psi - I(A : R)_\Psi - 5\eta(2\sqrt{\epsilon}) \log d_R \\
&= S(R)_\Psi + S(AR)_\Psi - S(A)_\Psi - 5\eta(2\sqrt{\epsilon}) \log d_R \\
&= S(R)_\Psi + S(B)_\Psi - S(BR)_\Psi - 5\eta(2\sqrt{\epsilon}) \log d_R \\
&= I(B : R)_\Psi - 5\eta(2\sqrt{\epsilon}) \log d_R,
\end{aligned}$$

which concludes the proof.  $\blacksquare$

## APPENDIX C

### PROOF OF LEMMA 18, 19 AND 20

#### A. Proof of Lemma 18

We prove that an  $M$ -induced map is  $(3\varsigma + 2\nu)$ -decoupling between  $A'R_A$  and  $R_B$  if it is  $\varsigma$ -oblivious and  $\nu$ -Markovianizing from  $R_A B R_B$ , which implies Lemma 18.

Due to Equalities (58), (68), (25) and (21), we have

$$\begin{aligned}
& \left\| \Phi_M^{R_A} - \pi_d^{R_A} \right\|_1 \\
&= \left\| \Phi_M^{R_A} \otimes \Phi_d^{B R_B} - \pi_d^{R_A} \otimes \Phi_d^{B R_B} \right\|_1 \\
&= \left\| \hat{U}^{R_A R_B} (\Phi_M^{R_A} \otimes \Phi_d^{B R_B}) \hat{U}^{\dagger R_A R_B} \right. \\
&\quad \left. - \hat{U}^{R_A R_B} (\pi_d^{R_A} \otimes \Phi_d^{B R_B}) \hat{U}^{\dagger R_A R_B} \right\|_1 \\
&= \left\| \Psi_M^{R_A B R_B} - \Psi^{R_A B R_B} \right\|_1.
\end{aligned}$$

Thus the condition of  $\varsigma$ -obliviousness is equivalent to

$$\left\| \Psi_M^{R_A B R_B} - \Psi^{R_A B R_B} \right\|_1 \leq \varsigma,$$

which implies

$$\left\| \Psi_M^{R_A R_B} - \pi_d^{R_A} \otimes \pi_d^{R_B} \right\|_1 \leq \varsigma$$

and

$$\left\| \Psi_M^{R_A} - \pi_d^{R_A} \right\|_1 \leq \varsigma, \quad \left\| \Psi_M^{R_B} - \pi_d^{R_B} \right\|_1 \leq \varsigma, \quad (95)$$

due to (66). From (56) and (58), it follows that

$$\begin{aligned}
& \left\| \Psi_M^{R_A} \otimes \Psi_M^{R_B} - \pi_d^{R_A} \otimes \pi_d^{R_B} \right\|_1 \\
&\leq \left\| \Psi_M^{R_A} \otimes \Psi_M^{R_B} - \pi_d^{R_A} \otimes \Psi_M^{R_B} \right\|_1 \\
&\quad + \left\| \pi_d^{R_A} \otimes \Psi_M^{R_B} - \pi_d^{R_A} \otimes \pi_d^{R_B} \right\|_1 \\
&= \left\| \Psi_M^{R_A} - \pi_d^{R_A} \right\|_1 + \left\| \Psi_M^{R_B} - \pi_d^{R_B} \right\|_1 \\
&\leq 2\varsigma,
\end{aligned}$$

and that

$$\begin{aligned}
& \left\| \Psi_M^{R_A R_B} - \Psi_M^{R_A} \otimes \Psi_M^{R_B} \right\|_1 \\
&\leq \left\| \Psi_M^{R_A R_B} - \pi_d^{R_A} \otimes \pi_d^{R_B} \right\|_1 \\
&\quad + \left\| \Psi_M^{R_A} \otimes \Psi_M^{R_B} - \pi_d^{R_A} \otimes \pi_d^{R_B} \right\|_1 \\
&\leq 3\varsigma. \quad (96)
\end{aligned}$$

Suppose  $\Psi_M^{A'R_A(BR_B)}$  is  $\nu$ -recoverable from  $R_A BR_B$ . By definition, there exists a linear CPTP map  $\mathcal{R} : R_A \rightarrow A'R_A$  such that

$$\left\| \Psi_M^{A'R_A(BR_B)} - \mathcal{R}(\Psi_M^{R_A BR_B}) \right\|_1 \leq \nu,$$

which implies

$$\left\| \Psi_M^{A'R_A} - \mathcal{R}(\Psi_M^{R_A}) \right\|_1 \leq \left\| \Psi_M^{A'R_A R_B} - \mathcal{R}(\Psi_M^{R_A R_B}) \right\|_1 \leq \nu. \quad (97)$$

Due to (65), Inequality (96) implies

$$\left\| \mathcal{R}(\Psi_M^{R_A R_B}) - \mathcal{R}(\Psi_M^{R_A}) \otimes \Psi_M^{R_B} \right\|_1 \leq 3\varsigma. \quad (98)$$

From (56), (58), (97) and (98), we obtain

$$\begin{aligned} & \left\| \Psi_M^{A'R_A R_B} - \Psi_M^{A'R_A} \otimes \Psi_M^{R_B} \right\|_1 \\ & \leq \left\| \Psi_M^{A'R_A R_B} - \mathcal{R}(\Psi_M^{R_A R_B}) \right\|_1 \\ & \quad + \left\| \mathcal{R}(\Psi_M^{R_A R_B}) - \mathcal{R}(\Psi_M^{R_A}) \otimes \Psi_M^{R_B} \right\|_1 \\ & \quad + \left\| \mathcal{R}(\Psi_M^{R_A}) \otimes \Psi_M^{R_B} - \Psi_M^{A'R_A} \otimes \Psi_M^{R_B} \right\|_1 \\ & \leq 3\varsigma + 2\nu, \end{aligned}$$

which completes the proof.  $\blacksquare$

### B. Proof of Lemma 19

We prove that an  $M$ -induced map is  $(\varsigma + \mu)$ -Markovianizing from  $A'R_A$  if it is  $\varsigma$ -oblivious and  $\mu$ -decoupling between  $A'R_A$  and  $R_B$ , which implies Lemma 19. Let  $\Xi : R_A R_B \rightarrow R_A BR_B$  be a linear CPTP map defined by

$$\Xi(\tau) = d^2 \cdot (\Psi^{R_A BR_B})^{\frac{1}{2}} (\tau^{R_A R_B} \otimes I^B) (\Psi^{R_A BR_B})^{\frac{1}{2}}.$$

This is indeed CPTP since we have

$$\begin{aligned} \text{Tr}[\Xi(\tau)] &= d^2 \cdot \text{Tr}[\tau^{R_A R_B} (\Psi^{R_A BR_B})] \\ &= d^2 \cdot \text{Tr}[\tau^{R_A R_B} (\pi_d^{R_A} \otimes \pi_d^{R_B})] \\ &= \text{Tr}[\tau]. \end{aligned}$$

Using the relation

$$\begin{aligned} & (\Psi^{R_A BR_B})^{\frac{1}{2}} \\ &= \left( U^{*R_A R_B} (\pi_d^{R_A} \otimes |\Phi_d\rangle\langle\Phi_d|^{BR_B}) U^{tR_A R_B} \right)^{\frac{1}{2}} \\ &= \frac{1}{\sqrt{d}} U^{*R_A R_B} (I^{R_A} \otimes |\Phi_d\rangle\langle\Phi_d|^{BR_B}) U^{tR_A R_B}, \end{aligned}$$

we have

$$\Xi^{R_A R_B}(\Psi^{A'R_A R_B}) = \Psi^{A'R_A BR_B},$$

which leads to

$$\Psi_M^{A'R_A BR_B} = \Xi^{R_A R_B}(\Psi_M^{A'R_A R_B}).$$

It is straightforward to verify that a map  $\mathcal{R}' : R_A \rightarrow R_A BR_B$  defined by

$$\mathcal{R}'(\tau) = \Xi(\tau^{R_A} \otimes \pi_d^{R_B}) \quad (\forall \tau \in \mathcal{S}(\mathcal{H}^{R_A}))$$

is CPTP as well. Therefore, from (65), (56) and (58), we have

$$\begin{aligned} & \left\| \Psi_M^{A'R_A BR_B} - \mathcal{R}'(\Psi_M^{A'R_A}) \right\|_1 \\ &= \left\| \Xi^{R_A R_B}(\Psi_M^{A'R_A R_B}) - \Xi^{R_A R_B}(\Psi_M^{A'R_A} \otimes \pi_d^{R_B}) \right\|_1 \\ &\leq \left\| \Psi_M^{A'R_A R_B} - \Psi_M^{A'R_A} \otimes \pi_d^{R_B} \right\|_1 \\ &\leq \left\| \Psi_M^{A'R_A R_B} - \Psi_M^{A'R_A} \otimes \Psi_M^{R_B} \right\|_1 \\ &\quad + \left\| \Psi_M^{A'R_A} \otimes \Psi_M^{R_B} - \Psi_M^{A'R_A} \otimes \pi_d^{R_B} \right\|_1 \\ &= \left\| \Psi_M^{A'R_A R_B} - \Psi_M^{A'R_A} \otimes \Psi_M^{R_B} \right\|_1 + \left\| \Psi_M^{R_B} - \pi_d^{R_B} \right\|_1 \\ &\leq \varsigma + \mu, \end{aligned}$$

where the last line follows from the assumption and (95).  $\blacksquare$

### C. Proof of Lemma 20

Suppose that an  $M_k$ -induced map is  $\mu_k$ -decoupling between  $A'R_A$  and  $R_B$  for each  $k \in \mathbb{K}$ , and that  $\sum_{k \in \mathbb{K}} p_k \mu_k \leq \mu$ . Due to (24) and Equality (62), there exist pure states  $|\Psi_k^p\rangle^{A'R_A \tilde{B}}$ ,  $|\Psi_k^q\rangle^{BR_B}$  and isometries  $W_k^{B B_0 \rightarrow B \tilde{B}}$  ( $k \in \mathbb{K}$ ) such that

$$\begin{aligned} & \left\| \Psi'_k - (\Psi_k^p)^{A'R_A \tilde{B}} \otimes (\Psi_k^q)^{BR_B} \right\|_1 \leq 2\sqrt{\mu_k}, \\ & (\Psi_k^p)^{A'R_A} = \Psi_{M_k}^{A'R_A}, \quad (\Psi_k^q)^{R_B} = \Psi_{M_k}^{R_B}, \end{aligned} \quad (99)$$

where  $|\Psi_k'\rangle := W_k |\Psi_{M_k}\rangle$ . Suppose in addition that  $\mathbb{M}$  is 0-oblivious. Then we have

$$(\Psi_k^p)^{R_A} = \pi_d^{R_A}, \quad (\Psi_k^q)^{R_B} = \pi_d^{R_B} \quad (100)$$

for each  $k$ , due to (95) and (99). The latter of (100) implies we can choose  $|\Psi_k^q\rangle^{BR_B} = |\Phi_d\rangle^{BR_B}$  by an appropriate choice of  $W_k$ , since all purifications are local-isometry equivalent. If  $\Psi_{M_k}^{A'R_A R_B}$  does not depend on  $k$ , neither does  $\Psi_{M_k}^{A'R_A}$ . Thus the  $k$ -dependence of  $|\Psi_k'\rangle$  and  $|\Psi_k^p\rangle$  can be dropped by an appropriate choice of  $W_k$  for the same reason. Hence we obtain

$$\left\| |\Psi'\rangle\langle\Psi'| - (\Psi^p)^{A'R_A \tilde{B}} \otimes \Phi_d^{BR_B} \right\|_1 \leq 2\sqrt{\mu_k},$$

for any  $k \in \mathbb{K}$ , which leads to

$$\left\| |\Psi'\rangle\langle\Psi'| - (\Psi^p)^{A'R_A \tilde{B}} \otimes \Phi_d^{BR_B} \right\|_1 \leq 2\sqrt{\mu} \quad (101)$$

and

$$\left\| (\Psi')^{\tilde{B} B} - (\Psi^p)^{\tilde{B}} \otimes \pi_d^B \right\|_1 \leq 2\sqrt{\mu} \quad (102)$$

due to (66).

Let  $\{\lambda_i^\downarrow\}_{i=1}^{\dim \tilde{B}}$  be the eigenvalues of  $(\Psi^p)^{\tilde{B}}$  sorted in decreasing order, let

$$(\Psi^p)^{\tilde{B}} = \sum_{i=1}^{\dim \tilde{B}} \lambda_i^\downarrow |i\rangle\langle i|$$

be the eigen decomposition of  $(\Psi^p)^{\tilde{B}}$ , and define a linear operator  $\Pi$  on  $\mathcal{H}^{\tilde{B}}$  by

$$\Pi := \sum_{i=1}^{\dim B_0} |i\rangle\langle i|.$$

II-7. Alice discards  $A_E$ .

*Remark.* In the description of  $\mathcal{M}$  by I-1~5, we assume that all information about the outcome of Alice's measurement, represented by  $k \in \mathbb{K}$ , is communicated to Bob. In a general protocol, however, not all information about the measurement outcome need to be communicated. In such cases, the measurement outcomes are represented as  $(k_1, k_2) \in \mathbb{K}_1 \times \mathbb{K}_2$  by two countable sets  $\mathbb{K}_1$  and  $\mathbb{K}_2$ . The  $\mathbb{K}_1$  part of the outcome is communicated to Bob, whereas the  $\mathbb{K}_2$  part is kept on Alice's register until she performs the last operation. We show that such protocols can also be described by II-1~7 as follows. Let  $\tilde{V} : A \rightarrow A' A_{E_1} A_{E_2}$  be an isometry such that the Naimark extension of Alice's measurement is given by  $M_{k_1 k_2} = \langle k_1 |^{A_{E_1}} \langle k_2 |^{A_{E_2}} \tilde{V}$ , and let  $V_{k_1 k_2 l} : A' \rightarrow A A_1 A_E$  be an isometry such that the Stinespring delation of Alice's last operation is given by  $\mathcal{O}_{k_1 k_2 l}(\tau) = \text{Tr}_{A_E} [V_{k_1 k_2 l} \tau^{A'} V_{k_1 k_2 l}^\dagger]$ . The procedure II-1~7 then gives a description of the general protocol by the following correspondence:

$$\begin{aligned} k &\rightarrow k_1 \\ A' &\rightarrow A' A_{E_2} \\ M_k^{A A_0 \rightarrow A'} &\rightarrow M_{k_1}^{A A_0 \rightarrow A' A_{E_2}} = \langle k_1 |^{A_{E_1}} \tilde{V} \\ V_{kl} &\rightarrow V_{k_1 l} = \sum_{k_2 \in \mathbb{K}_2} |k_2\rangle \langle k_2|^{A_{E_2}} \otimes V_{k_1 k_2 l} \\ A_E &\rightarrow A_{E_2} A_E. \end{aligned}$$

### B. Proof of Lemma 22

We prove that the measurement  $\mathbb{M}$  is  $4\sqrt[4]{\epsilon}$ -oblivious and  $8\sqrt[4]{\epsilon}$ -decoupling between  $A'R_A$  and  $R_B$ , which implies Lemma 22 combined with Lemma 19. From (22) and Equality (64), we have

$$\begin{aligned} 1 - \epsilon &\leq \sum_{kl} p_{kl} F\left(\hat{\Psi}_{kl}^{ABR_A R_B}, |\Phi_d\rangle^{A R_A} |\Phi_d\rangle^{B R_B}\right) \\ &= \sum_{kl} p_{kl} F\left(|\hat{\Psi}_{kl}\rangle, |\Phi_d\rangle^{A R_A} |\Phi_d\rangle^{B R_B} |\phi_{kl}\rangle^{A_1 B_1 A_E}\right) \end{aligned}$$

for some states  $\phi_{kl}$ , which leads to

$$\sum_{kl} p_{kl} \epsilon_{kl} \leq \epsilon$$

for

$$\epsilon_{kl} := 1 - F\left(|\hat{\Psi}_{kl}\rangle, |\Phi_d\rangle^{A R_A} |\Phi_d\rangle^{B R_B} |\phi_{kl}\rangle^{A_1 B_1 A_E}\right). \quad (108)$$

Due to Lemma 34 and  $\epsilon_{kl} \in [0, 1]$ , we have

$$\sum_{kl} p_{kl} \sqrt{\epsilon_{kl}} \leq \sqrt[4]{\epsilon} + \sqrt{\epsilon} \leq 2\sqrt[4]{\epsilon}. \quad (109)$$

Therefore, by using (70), (66) and (57), we obtain

$$\begin{aligned} 4\sqrt[4]{\epsilon} &\geq 2 \sum_{kl} p_{kl} \sqrt{\epsilon_{kl}} \\ &\geq \sum_{kl} p_{kl} \left\| |\hat{\Psi}_{kl}\rangle \langle \hat{\Psi}_{kl}| - \Phi_d^{A R_A} \otimes \Phi_d^{B R_B} \otimes \phi_{kl}^{A_1 B_1 A_E} \right\|_1 \\ &\geq \sum_{kl} p_{kl} \left\| \hat{\Psi}_{kl}^{A R_A R_B A_1 A_E} - \Phi_d^{A R_A} \otimes \pi_d^{R_B} \otimes \phi_{kl}^{A_1 A_E} \right\|_1 \end{aligned}$$

$$\begin{aligned} &= \sum_{kl} p_{kl} \left\| \Psi_{kl}^{A' R_A R_B} - V_{kl}^\dagger (\Phi_d^{A R_A} \otimes \phi_{kl}^{A_1 A_E}) V_{kl} \otimes \pi_d^{R_B} \right\|_1 \\ &\geq \sum_k p_k \left\| \sum_l p_{l|k} \Psi_{kl}^{A' R_A R_B} - \psi_k^{A' R_A} \otimes \pi_d^{R_B} \right\|_1 \\ &= \sum_k p_k \left\| \Psi_k^{A' R_A R_B} - \psi_k^{A' R_A} \otimes \pi_d^{R_B} \right\|_1 \\ &= \sum_k p_k \left\| \Psi_k^{A' R_A R_B} - \psi_k^{A' R_A} \otimes \pi_d^{R_B} \right\|_1, \end{aligned} \quad (110)$$

where we defined

$$\psi_k^{A' R_A} := \sum_l p_{l|k} V_{kl}^\dagger (\Phi_d^{A R_A} \otimes \phi_{kl}^{A_1 A_E}) V_{kl}. \quad (111)$$

Hence, from (56), (58) and (65), we obtain

$$\begin{aligned} &\sum_k p_k \left\| \Psi_k^{A' R_A R_B} - \Psi_k^{A' R_A} \otimes \Psi_k^{R_B} \right\|_1 \\ &\leq \sum_k p_k \left\| \Psi_k^{A' R_A R_B} - \psi_k^{A' R_A} \otimes \pi_d^{R_B} \right\|_1 \\ &\quad + \sum_k p_k \left\| \psi_k^{A' R_A} \otimes \pi_d^{R_B} - \Psi_k^{A' R_A} \otimes \pi_d^{R_B} \right\|_1 \\ &\quad + \sum_k p_k \left\| \Psi_k^{A' R_A} \otimes \pi_d^{R_B} - \Psi_k^{A' R_A} \otimes \Psi_k^{R_B} \right\|_1 \\ &= \sum_k p_k \left\| \Psi_k^{A' R_A R_B} - \psi_k^{A' R_A} \otimes \pi_d^{R_B} \right\|_1 \\ &\quad + \sum_k p_k \left\| \psi_k^{A' R_A} - \Psi_k^{A' R_A} \right\|_1 + \sum_k p_k \left\| \pi_d^{R_B} - \Psi_k^{R_B} \right\|_1 \\ &\leq 3 \sum_k p_k \left\| \Psi_k^{A' R_A R_B} - \psi_k^{A' R_A} \otimes \pi_d^{R_B} \right\|_1 \leq 12\sqrt[4]{\epsilon}, \end{aligned}$$

Thus Alice's measurement is  $12\sqrt[4]{\epsilon}$ -decoupling between  $A'R_A$  and  $R_B$ . From (110), (66), (111), (68), (25), (26) and (58), we also have

$$\begin{aligned} 4\sqrt[4]{\epsilon} &\geq \sum_k p_k \left\| \Psi_k^{R_A R_B} - \psi_k^{R_A} \otimes \pi_d^{R_B} \right\|_1 \\ &= \sum_k p_k \left\| \Psi_k^{R_A R_B} - \pi_d^{R_A} \otimes \pi_d^{R_B} \right\|_1 \\ &= \sum_k p_k \left\| \hat{U}^{R_A R_B} (\Phi_{M_k}^{R_A} \otimes \pi_d^{R_B}) \hat{U}^{\dagger R_A R_B} \right. \\ &\quad \left. - \hat{U}^{R_A R_B} (\pi_d^{R_A} \otimes \pi_d^{R_B}) \hat{U}^{\dagger R_A R_B} \right\|_1 \\ &= \sum_k p_k \left\| \Phi_{M_k}^{R_A} \otimes \pi_d^{R_B} - \pi_d^{R_A} \otimes \pi_d^{R_B} \right\|_1 \\ &= \sum_k p_k \left\| \Phi_{M_k}^{R_A} - \pi_d^{R_A} \right\|_1, \end{aligned}$$

which implies that Alice's measurement is  $4\sqrt[4]{\epsilon}$ -oblivious. ■

### C. Proof of Lemma 23

From (108), (67) and (107), we have

$$1 - \epsilon_{kl} \leq F(\hat{\Psi}_{kl}^{B R_B}, \Phi_d^{B R_B}) = F(\Psi_{kl}^{B R_B}, \Phi_d^{B R_B}).$$

By using (61), we have

$$\begin{aligned} 1 - \sum_l p_{l|k} \epsilon_{kl} &\leq F\left(\sum_l p_{l|k} \Psi_{kl}^{BR_B}, \Phi_d^{BR_B}\right) \\ &= F(\Psi_k^{BR_B}, \Phi_d^{BR_B}) \end{aligned}$$

because of (106). Due to Equality (64), there exist pure states  $|\Psi_k^p\rangle^{A'R_A B_1 B_E}$  ( $k \in \mathbb{K}$ ) such that

$$1 - \sum_l p_{l|k} \epsilon_{kl} \leq F(|\Psi_k^p\rangle, (\Psi_k^p)^{A'R_A B_1 B_E} \otimes \Phi_d^{BR_B})$$

for each  $k$ , which leads to

$$2\sqrt{\sum_l p_{l|k} \epsilon_{kl}} \geq \left\| \Psi_k' - (\Psi_k^p)^{A'R_A B_1 B_E} \otimes \Phi_d^{BR_B} \right\|_1$$

due to (70). Thus we have

$$\begin{aligned} \sum_k p_k \left\| \Psi_k' - (\Psi_k^p)^{A'R_A B_1 B_E} \otimes \Phi_d^{BR_B} \right\|_1 \\ \leq 2 \sum_k p_k \sqrt{\sum_l p_{l|k} \epsilon_{kl}} \leq 2 \sum_{kl} p_{kl} \sqrt{\epsilon_{kl}} \leq 4\sqrt[4]{\epsilon}, \end{aligned}$$

where the last line follows from the concavity of the square root function and Inequality (109). This completes the proof of Lemma 23.  $\blacksquare$

#### APPENDIX E PROOF OF INEQUALITY (33)

In this Appendix, we describe an evaluation of the total error, which have appeared in Section VI-A in the proof of the direct part of Theorem 5.

From Lemma 21, we have

$$\left\| |\Psi'\rangle\langle\Psi'| - (\tilde{\Psi}^p)^{\bar{A}\bar{R}_A\bar{B}} \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right\|_1 \leq 5\sqrt[4]{2\epsilon} \quad (112)$$

and

$$\left\| (\tilde{\Psi}^p)^{\bar{R}_A} - (\pi_d^{\otimes n})^{\bar{R}_A} \right\|_1 \leq 3\sqrt[4]{2\epsilon}, \quad (113)$$

corresponding to (30) and (31), respectively. Let  $A_{B'}$  be Alice's register which is identical to  $B'$ , and  $\mathcal{N} : A'B' \rightarrow A'A_{B'}$  be state merging of  $|\tilde{\Psi}^p\rangle^{A'\bar{R}_A B'}$ . Define the merging error  $\epsilon_{\text{merg}}$  by

$$\epsilon_{\text{merg}} := \left\| \mathcal{N}((\tilde{\Psi}^p)^{A'\bar{R}_A B'}) - (\tilde{\Psi}^p)^{A'\bar{R}_A A_{B'}} \right\|_1. \quad (114)$$

From (112) and (32), we have

$$\left\| |\Psi'\rangle\langle\Psi'| \otimes \Phi_{2^{nr}}^{\bar{A}_0\bar{B}_0} - (\tilde{\Psi}^p)^{A'\bar{R}_A B'} \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right\|_1 \leq 5\sqrt[4]{2\epsilon},$$

which leads to

$$\begin{aligned} \left\| \mathcal{N}\left(|\Psi'\rangle\langle\Psi'| \otimes \Phi_{2^{nr}}^{\bar{A}_0\bar{B}_0}\right) - \right. \\ \left. \mathcal{N}((\tilde{\Psi}^p)^{A'\bar{R}_A B'}) \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right\|_1 \leq 5\sqrt[4]{2\epsilon}, \quad (115) \end{aligned}$$

By definition, we have  $(\tilde{\Psi}^p)^{\bar{R}_A} = (\tilde{\Psi}^p)^{\bar{R}_A}$ . Therefore, due to Inequality (113) and Lemma 31, there exists a quantum operation  $\mathcal{O} : A'A_{B'} \rightarrow \bar{A}$  such that

$$\left\| \mathcal{O}^{A'A_{B'}}((\tilde{\Psi}^p)^{A'\bar{R}_A A_{B'}}) - (\Phi_d^{\otimes n})^{\bar{A}\bar{R}_A} \right\|_1 \leq 2\sqrt[3]{3}\sqrt[8]{2\epsilon}. \quad (116)$$

Define a quantum operation  $\mathcal{N}' : A'B' \rightarrow \bar{A}$  by  $\mathcal{N}' := \mathcal{O} \circ \mathcal{N}$ . From (114), (115) and (65), we obtain

$$\left\| \mathcal{N}'((\tilde{\Psi}^p)^{A'\bar{R}_A B'}) - \mathcal{O}^{A'A_{B'}}((\tilde{\Psi}^p)^{A'\bar{R}_A A_{B'}}) \right\|_1 \leq \epsilon_{\text{merg}}, \quad (117)$$

$$\begin{aligned} \left\| \mathcal{N}'\left(|\Psi'\rangle\langle\Psi'| \otimes \Phi_{2^{nr}}^{\bar{A}_0\bar{B}_0}\right) - \right. \\ \left. \mathcal{N}'((\tilde{\Psi}^p)^{A'\bar{R}_A B'}) \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right\|_1 \leq 5\sqrt[4]{2\epsilon}. \quad (118) \end{aligned}$$

From (56), (58), (116), (117) and (118), we see that

$$\begin{aligned} &\left\| \mathcal{N}'\left(|\Psi'\rangle\langle\Psi'| \otimes \Phi_{2^{nr}}^{\bar{A}_0\bar{B}_0}\right) - (\Phi_d^{\otimes n})^{\bar{A}\bar{R}_A} \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right\|_1 \\ &\leq \left\| \mathcal{N}'\left(|\Psi'\rangle\langle\Psi'| \otimes \Phi_{2^{nr}}^{\bar{A}_0\bar{B}_0}\right) - \right. \\ &\quad \left. \mathcal{N}'((\tilde{\Psi}^p)^{A'\bar{R}_A B'}) \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right\|_1 \\ &\quad + \left\| \mathcal{N}'((\tilde{\Psi}^p)^{A'\bar{R}_A B'}) \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} - \right. \\ &\quad \left. \mathcal{O}^{A'A_{B'}}((\tilde{\Psi}^p)^{A'\bar{R}_A A_{B'}}) \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right\|_1 \\ &\quad + \left\| \mathcal{O}^{A'A_{B'}}((\tilde{\Psi}^p)^{A'\bar{R}_A A_{B'}}) \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} - \right. \\ &\quad \left. (\Phi_d^{\otimes n})^{\bar{A}\bar{R}_A} \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right\|_1 \\ &\leq \left\| \mathcal{N}'\left(|\Psi'\rangle\langle\Psi'| \otimes \Phi_{2^{nr}}^{\bar{A}_0\bar{B}_0}\right) - \right. \\ &\quad \left. \mathcal{N}'((\tilde{\Psi}^p)^{A'\bar{R}_A B'}) \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right\|_1 \\ &\quad + \left\| \mathcal{N}'((\tilde{\Psi}^p)^{A'\bar{R}_A B'}) - \mathcal{O}^{A'A_{B'}}((\tilde{\Psi}^p)^{A'\bar{R}_A A_{B'}}) \right\|_1 \\ &\quad + \left\| \mathcal{O}^{A'A_{B'}}((\tilde{\Psi}^p)^{A'\bar{R}_A A_{B'}}) - (\Phi_d^{\otimes n})^{\bar{A}\bar{R}_A} \right\|_1 \\ &\leq 2\sqrt[3]{3}\sqrt[8]{2\epsilon} + 5\sqrt[4]{2\epsilon} + \epsilon_{\text{merg}}. \end{aligned}$$

Due to (17), (18) and (70), we have

$$\epsilon_{\text{merg}} \leq 2 \left( 2\sqrt{2}\sqrt{2^{-\frac{nr}{2}} + 2^{-n(R+r)}} \right)^{\frac{1}{2}} \leq 4 \cdot 2^{-nr/8}.$$

Thus we obtain (33).  $\blacksquare$

#### APPENDIX F PROOF OF THE CONVERSE PART

Suppose a rate triplet  $(E, C_f, C_b)$  is achievable by a two-round protocol. By Definition 2 and Assumption (7), for any  $\delta \in (0, 1]$ , there exist  $\epsilon > 0$ ,  $n$  satisfying

$$8\sqrt[4]{\epsilon} \cdot n \leq \delta \quad (119)$$

and a  $(2, n, \epsilon)$ -protocol  $\mathcal{M}_n$  for implementing  $U$  with the entanglement cost  $nE$ , the classical communication cost  $nC_f$  and the backward classical communication cost  $nC_b$ . We assume here for simplicity that  $K_n$  and  $L_n$  is bounded above as

$$\log K_n, \log L_n \leq n \log \kappa \quad (120)$$

with a constant  $\kappa > 0$ . As we prove below, the following inequalities hold for any such  $\mathcal{M}_n$ :

$$C_f \geq M(U^\dagger) - \tilde{\xi}(\delta) \log d, \quad (121)$$

$$\begin{aligned} n \log d + \log K_n - \sum_{k \in \mathbb{K}} p_k S(A')_{\Psi_k} \\ \geq nM(U^\dagger) - n\tilde{\xi}(\delta) \log d, \end{aligned} \quad (122)$$

$$\frac{1}{n}(\log K_n - \log L_n) \geq M(U^\dagger) - \xi_1(\delta) \log(d\kappa), \quad (123)$$

$$C_b \geq M(U^\dagger) - \xi_2(\delta) \log(d\kappa \cdot 2^{C_b}). \quad (124)$$

Here,  $\tilde{\xi}$  is a function defined by (87), and  $\xi_1, \xi_2$  are nonnegative functions that are independent of  $n$  and  $d$ , and satisfy  $\lim_{\delta \rightarrow 0} \xi_1(\delta) = \lim_{\delta \rightarrow 0} \xi_2(\delta) = 0$ . The converse part of Theorem 5 immediately follows by taking the limit of  $\delta \rightarrow 0$  in Inequalities (121), (123) and (124).

Let us prove Inequalities (121)~(124). From Lemma 22, Alice's measurement in  $\mathcal{M}_n$  is  $4\sqrt[4]{\epsilon}$ -oblivious and  $12\sqrt[4]{\epsilon}$ -Markovianizing from  $A'R_A$ . Hence Conditions 1) and 2) in Definition 9 are satisfied by the correspondence described by (35). Thus we can apply Lemma 11 to obtain the above four inequalities.

Inequality (121) follows from  $nC_f \geq H(\{p_k\}_{k \in \mathbb{K}})$  and Inequality (16). Inequality (122) follows from Inequality (16) on  $\Delta S(A')_{av}$ . We prove Inequalities (123) and (124) in the following subsections.

#### A. Proof of Inequality (123)

Let  $\tilde{\mathcal{M}}_{n,k} : A'B_1B_E \rightarrow \bar{A}A_1B_1$  be a CPTP map that describes the procedure II-4~7, presented in Appendix D-A, averaged over the measurement outcome  $l$ . The final state is given by

$$\rho(\mathcal{M}_n, U^\dagger) = \sum_k p_k \tilde{\mathcal{M}}_{n,k}(\Psi'_k). \quad (125)$$

Define

$$\epsilon_{k,1} := \left\| \tilde{\mathcal{M}}_{n,k}(\Psi'_k) - (\Phi_d^{\otimes n})^{\bar{A}\bar{R}_A} \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \otimes \Phi_{L_n}^{A_1B_1} \right\|_1. \quad (126)$$

and

$$f_{k,1} := F(\tilde{\mathcal{M}}_{n,k}(\Psi'_k), |\Phi_d^{\otimes n}\rangle^{\bar{A}\bar{R}_A} |\Phi_d^{\otimes n}\rangle^{\bar{B}\bar{R}_B} |\Phi_{L_n}\rangle^{A_1B_1})$$

for  $k \in \mathbb{K}$ . Due to the convexity of the square function, Inequality (70), Equalities (61), (125) and Inequality (34), we have

$$\begin{aligned} \left( \sum_k p_k \epsilon_{k,1} \right)^2 &\leq \sum_k p_k \epsilon_{k,1}^2 \leq 4 \sum_k p_k (1 - f_{k,1}) \\ &= 4 - 4F(\rho(\mathcal{M}_n, U^\dagger), |\Phi_d^{\otimes n}\rangle^{\bar{A}\bar{R}_A} |\Phi_d^{\otimes n}\rangle^{\bar{B}\bar{R}_B} |\Phi_{L_n}\rangle^{A_1B_1}) \\ &\leq 4\epsilon, \end{aligned}$$

which yields

$$\sum_k p_k \epsilon_{k,1} \leq 2\sqrt{\epsilon}. \quad (127)$$

From Lemma 23, there exist pure states  $|\Psi_k^p\rangle^{A'\bar{R}_AB_1B_E}$  such that

$$\sum_k p_k \left\| \Psi'_k - (\Psi_k^p)^{A'\bar{R}_AB_1B_E} \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right\|_1 \leq 4\sqrt[4]{\epsilon}.$$

Defining

$$\epsilon_{k,2} := \left\| \Psi'_k - (\Psi_k^p)^{A'\bar{R}_AB_1B_E} \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right\|_1, \quad (128)$$

we obtain

$$\sum_k p_k \epsilon_{k,2} \leq 4\sqrt[4]{\epsilon}. \quad (129)$$

From (127) and (129), we have

$$\sum_k p_k \epsilon_k \leq 2\sqrt{\epsilon} + 4\sqrt[4]{\epsilon} \leq 6\sqrt[4]{\epsilon} \quad (130)$$

for

$$\epsilon_k := \epsilon_{k,1} + \epsilon_{k,2}. \quad (131)$$

From (128) and (65), we have

$$\left\| \tilde{\mathcal{M}}_{n,k}(\Psi'_k) - \tilde{\mathcal{M}}_{n,k}(\Psi_k^p)^{\bar{A}\bar{R}_AA_1B_1} \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right\|_1 \leq \epsilon_{k,2} \quad (132)$$

for each  $k$ , which implies

$$\left\| (\Psi_k)^{A'} - (\Psi_k^p)^{A'} \right\|_1 = \left\| (\Psi'_k)^{A'} - (\Psi_k^p)^{A'} \right\|_1 \leq \epsilon_{k,2} \leq \epsilon_k \quad (133)$$

due to (66) and (105). By (58), (56), (132), (126) and (131), we see that

$$\begin{aligned} &\left\| (\tilde{\mathcal{M}}_{n,k}(\Psi_k^p))^{\bar{A}\bar{R}_AA_1B_1} - (\Phi_d^{\otimes n})^{\bar{A}\bar{R}_A} \otimes \Phi_{L_n}^{A_1B_1} \right\|_1 \\ &= \left\| (\tilde{\mathcal{M}}_{n,k}(\Psi_k^p))^{\bar{A}\bar{R}_AA_1B_1} \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right. \\ &\quad \left. - (\Phi_d^{\otimes n})^{\bar{A}\bar{R}_A} \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \otimes \Phi_{L_n}^{A_1B_1} \right\|_1 \\ &\leq \left\| \tilde{\mathcal{M}}_{n,k}(\Psi'_k) - \tilde{\mathcal{M}}_{n,k}(\Psi_k^p)^{\bar{A}\bar{R}_AA_1B_1} \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \right\|_1 \\ &\quad + \left\| \tilde{\mathcal{M}}_{n,k}(\Psi'_k) - (\Phi_d^{\otimes n})^{\bar{A}\bar{R}_A} \otimes (\Phi_d^{\otimes n})^{\bar{B}\bar{R}_B} \otimes \Phi_{L_n}^{A_1B_1} \right\|_1 \\ &\leq \epsilon_k, \end{aligned} \quad (134)$$

which leads to

$$\left\| (\Psi_k^p)^{\bar{R}_A} - (\pi_d^{\otimes n})^{\bar{R}_A} \right\|_1 \leq \epsilon_k. \quad (135)$$

by (66). Therefore, due to Lemma 31, there exists a quantum operation  $\mathcal{T}_{n,k} : \bar{A} \rightarrow A'B_{1B_E}$ , where  $A_{B_1B_E}$  is a quantum system which is identical to  $B_1B_E$ , such that

$$\left\| (\Psi_k^p)^{A'\bar{R}_AA_{B_1B_E}} - \mathcal{T}_{n,k}((\Phi_d^{\otimes n})^{\bar{A}\bar{R}_A}) \right\|_1 \leq 2\sqrt{\epsilon_k}. \quad (136)$$

Define

$$\tilde{\mathcal{M}}'_{n,k} := \mathcal{T}_{n,k} \circ \tilde{\mathcal{M}}_{n,k}. \quad (137)$$



Owing to (56), (137), (65), (58) and Inequalities (134), (136), we have

$$\begin{aligned}
& \left\| (\tilde{\mathcal{M}}'_{n,k}(\Psi_k^p))^{A' \bar{R}_A A_{B_1 B_E} A_1 B_1} \right. \\
& \quad \left. - (\Psi_k^p)^{A' \bar{R}_A A_{B_1 B_E}} \otimes \Phi_{L_n}^{A_1 B_1} \right\|_1 \\
& \leq \left\| (\tilde{\mathcal{M}}'_{n,k}(\Psi_k^p))^{A' \bar{R}_A A_{B_1 B_E} A_1 B_1} \right. \\
& \quad \left. - \mathcal{T}_{n,k}((\Phi_d^{\otimes n})^{\bar{A} \bar{R}_A}) \otimes \Phi_{L_n}^{A_1 B_1} \right\|_1 \\
& \quad + \left\| \mathcal{T}_{n,k}((\Phi_d^{\otimes n})^{\bar{A} \bar{R}_A}) \otimes \Phi_{L_n}^{A_1 B_1} \right. \\
& \quad \left. - (\Psi_k^p)^{A' \bar{R}_A A_{B_1 B_E}} \otimes \Phi_{L_n}^{A_1 B_1} \right\|_1 \\
& \leq \left\| (\tilde{\mathcal{M}}'_{n,k}(\Psi_k^p))^{\bar{A} \bar{R}_A A_1 B_1} - (\Phi_d^{\otimes n})^{\bar{A} \bar{R}_A} \otimes \Phi_{L_n}^{A_1 B_1} \right\|_1 \\
& \quad + \left\| \mathcal{T}_{n,k}((\Phi_d^{\otimes n})^{\bar{A} \bar{R}_A}) - (\Psi_k^p)^{A' \bar{R}_A A_{B_1 B_E}} \right\|_1 \\
& \leq \epsilon_k + 2\sqrt{\epsilon_k},
\end{aligned}$$

which leads to

$$\begin{aligned}
& F\left(\tilde{\mathcal{M}}'_{n,k}(\Psi_k^p), |\Psi_k^p\rangle^{A' \bar{R}_A A_{B_1 B_E}} |\Phi_{L_n}\rangle^{A_1 B_1}\right) \\
& \geq 1 - \min\{\epsilon_k + 2\sqrt{\epsilon_k}, 1\}.
\end{aligned}$$

Hence  $\tilde{\mathcal{M}}'_{n,k}$  is a state merging of  $|\Psi_k^p\rangle^{A' \bar{R}_A (B_1 B_E)}$  with the error  $\min\{\epsilon_k + 2\sqrt{\epsilon_k}, 1\}$  and the entanglement cost  $-\log L_n$  (see Definition 12 and Inequality (70)).

Note that we have

$$\dim \mathcal{H}^{A'} \leq \dim \mathcal{H}^{\bar{A}} \times \dim \mathcal{H}^{A_0} \leq d^n \times \kappa^n, \quad (138)$$

which corresponds to (23), and Assumption (120). Therefore, we apply Theorem 35, Inequalities (133), (135) and (78) to obtain

$$\begin{aligned}
& -\log L_n + n\eta' \left( 2\sqrt{\min\{\epsilon_k + 2\sqrt{\epsilon_k}, 1\}} \right) \log(d\kappa) \\
& \geq S(B_1 B_E | A')_{\Psi_k^p} \\
& = S(\bar{R}_A)_{\Psi_k^p} - S(A')_{\Psi_k^p} \\
& \geq n \log d - S(A')_{\Psi_k} - 2n\eta(\epsilon_k) \log(d\kappa)
\end{aligned}$$

for each  $k$ , where  $\eta'$  is a function defined by (88). Combining this with (139) and averaging over  $k$ , we obtain

$$\begin{aligned}
-\log L_n & \geq n \log d - \sum_k p_k S(A')_{\Psi_k} \\
& \quad - n \sum_k p_k \eta''(\epsilon_k) \log(d\kappa), \quad (139)
\end{aligned}$$

where we defined

$$\eta''(\epsilon) := \eta' \left( 2\sqrt{\min\{\epsilon + 2\sqrt{\epsilon}, 1\}} \right) + 2\eta(\epsilon). \quad (140)$$

Let us define a function  $\eta_1$  by

$$\eta_1(\epsilon) := \eta'' \left( \sqrt{6\sqrt[4]{\epsilon}} \right) + \eta''(1) \cdot \sqrt{6\sqrt[4]{\epsilon}}, \quad (141)$$

which satisfies  $\lim_{\epsilon \rightarrow 0} \eta_1(\epsilon) = 0$ . From Inequality (139), Lemma 34 and (130), we have

$$-\log L_n \geq n \log d - \sum_k p_k S(A')_{\Psi_k} - \eta_1(\epsilon) \log(d\kappa),$$

where we used the fact that we have  $6\sqrt[4]{\epsilon} \leq 1$  from (119). Combining this with Inequality (122), and defining  $\xi_1(\delta) := \tilde{\xi}(\delta) + \eta_1(\delta)$ , we obtain Inequality (123).

### B. Proof of Inequality (124)

To prove Inequality (124), note that  $\tilde{\mathcal{M}}'_{n,k}$  is a state merging of  $|\Psi_k^p\rangle^{A' \bar{R}_A (B_1 B_E)}$  with the error  $\min\{\epsilon_k + 2\sqrt{\epsilon_k}, 1\}$  and the classical communication cost  $nC_b$ . Therefore, from Theorem 35, Inequalities (135) and (78), we have

$$\begin{aligned}
& nC_b + 5n\eta \left( 2\sqrt{\min\{\epsilon_k + 2\sqrt{\epsilon_k}, 1\}} \right) \log(d\kappa) \\
& \geq I(B_1 B_E : \bar{R}_A)_{\Psi_k^p} \\
& = S(B_1 B_E)_{\Psi_k^p} + S(\bar{R}_A)_{\Psi_k^p} - S(B_1 B_E \bar{R}_A)_{\Psi_k^p} \\
& \geq S(B_1 B_E)_{\Psi_k^p} + n \log d - S(A')_{\Psi_k^p} - n\eta(\epsilon_k) \log d.
\end{aligned} \quad (142)$$

for each  $k$ . From (128), (78), (105), (131) and

$$\log \dim \mathcal{H}^{B_E} = nC_b, \quad \Psi_k^{B_0} = (\pi_d^{\otimes n})^{\bar{B}} \otimes \Psi_k^{B_0},$$

we have

$$\begin{aligned}
& S(B_1 B_E)_{\Psi_k^p} = S(\bar{B} B_1 B_E)_{\Psi_k^p \otimes \Phi_d^{\otimes n}} - S(\bar{B})_{\Phi_d^{\otimes n}} \\
& \geq S(\bar{B} B_1 B_E)_{\Psi_k^p} - n \log d - n\eta(\epsilon_{k,2}) \log(d\kappa \cdot 2^{C_b}) \\
& = S(\bar{B} B_0)_{\Psi_k} - n \log d - n\eta(\epsilon_{k,2}) \log(d\kappa \cdot 2^{C_b}) \\
& = S(\bar{B})_{\pi_d^{\otimes n}} + S(B_0)_{\Psi_k} - n \log d - n\eta(\epsilon_{k,2}) \log(d\kappa \cdot 2^{C_b}) \\
& = S(B_0)_{\Psi_k} - n\eta(\epsilon_{k,2}) \log(d\kappa \cdot 2^{C_b}) \\
& \geq S(B_0)_{\Psi_k} - n\eta(\epsilon_k) \log(d\kappa \cdot 2^{C_b}).
\end{aligned} \quad (143)$$

Using (128), (138), (105) and (131), we also have

$$\begin{aligned}
S(A')_{\Psi_k^p} & \leq S(A')_{\Psi_k^p} + n\eta(\epsilon_{k,2}) \log(d\kappa) \\
& \leq S(A')_{\Psi_k} + n\eta(\epsilon_k) \log(d\kappa).
\end{aligned} \quad (144)$$

From (142), (143), (144) and (140), we see that

$$\begin{aligned}
& nC_b \\
& \geq n \log d - S(A')_{\Psi_k} + S(B_0)_{\Psi_k} \\
& \quad - n \left( 5\eta \left( 2\sqrt{\min\{\epsilon + 2\sqrt{\epsilon}, 1\}} \right) + 3\eta(\epsilon_k) \right) \log(d\kappa \cdot 2^{C_b}) \\
& \geq n \log d - S(A')_{\Psi_k} + S(B_0)_{\Psi_k} - 5n\eta''(\epsilon_k) \log(d\kappa \cdot 2^{C_b})
\end{aligned}$$

for each  $k$ . Averaging over  $k$ , we obtain

$$\begin{aligned}
nC_b & \geq \sum_k p_k (n \log d - S(A')_{\Psi_k} + S(B_0)_{\Psi_k}) \\
& \quad - 5n\eta_1(\epsilon) \log(d\kappa \cdot 2^{C_b}),
\end{aligned}$$

where  $\eta_1$  is a function defined by (141). Thus, from Inequality (16) on  $\Delta S(A')_{av} - \Delta S(G)_{av}$ , we have

$$nC_b \geq nM(U^\dagger) - n(\tilde{\xi}(\delta) + 5\eta_1(\delta)) \log(d\kappa \cdot 2^{C_b}).$$

Defining  $\xi_2(\delta) := \tilde{\xi}(\delta) + 5\eta_1(\delta)$ , we obtain Inequality (124).  $\blacksquare$

### C. On the Convergence Speed of the Error

We prove that the converse part of Theorem 5 holds even when we drop Condition (7), if Conjecture 28 is true. First, as we proved in Appendix E of [14] (see Remark therein), the function  $\xi(\delta)$  in (86) can be replaced by another function  $\xi_{av}(\epsilon)$ , which is independent of  $n, \delta$  and satisfies  $\lim_{\epsilon \rightarrow 0} \xi_{av}(\epsilon) = 0$ . Consequently, functions  $\xi(\delta)$ ,  $\xi_1(\delta)$  and  $\xi_2(\delta)$  in Inequalities (121) ~ (124) can be replaced by different functions  $\xi'(\epsilon)$ ,  $\xi'_1(\epsilon)$  and  $\xi'_2(\epsilon)$ , respectively, which do not depend on  $n, \delta$  and vanishes in the limit of  $\epsilon \rightarrow 0$ . Inequalities (121) ~ (124) then hold for any  $n$  and  $\epsilon$ , which implies that the converse part holds without additional assumption (7).

## APPENDIX G PROOF OF THEOREM 26

We prove Theorem 26 after introducing a theorem regarding the cost of randomness for destroying correlations in a bipartite quantum state.

### A. Decoupling

The following lemma is obtained as a corollary of Proposition 2 in [39], except an evaluation of the convergence speed of the error.

*Lemma 36* Let  $\pi^A$  be the maximally mixed state on  $\mathcal{H}^A$ , and suppose a bipartite state  $\rho^{AB} \in \mathcal{S}(\mathcal{H}^A \otimes \mathcal{H}^B)$  satisfies  $\rho^A = \pi^A$ . There exists a constant  $c > 0$  that satisfies the following properties for any  $R > I(A : B)_\rho$ , sufficiently small  $\delta > 0$  and sufficiently large  $n$ . That is, for an arbitrary ensemble of unitaries on  $(\mathcal{H}^A)^{\otimes n}$  satisfying

$$\forall |\phi\rangle \in (\mathcal{H}^A)^{\otimes n}; \int_V p(dV) V |\phi\rangle \langle \phi| V^\dagger = (\pi^A)^{\otimes n}, \quad (145)$$

there exists a set of unitaries  $\{V_k\}_{k=1}^{2^{nR}}$  on the support of  $p(dV)$ , such that a random unitary operation  $\mathcal{V}_n$  on  $\mathcal{S}(\mathcal{H}^A)$  defined by

$$\mathcal{V}_n(\cdot) = \frac{1}{2^{nR}} \sum_{k=1}^{2^{nR}} V_k(\cdot) V_k^\dagger \quad (146)$$

satisfies

$$\begin{aligned} & \|\mathcal{V}_n((\rho^{AB})^{\otimes n}) - (\pi^A)^{\otimes n} \otimes (\rho^B)^{\otimes n}\|_1 \\ & \leq 2^{-n\delta} + 14 \exp\left(-\frac{c\delta^2 n}{4}\right). \end{aligned} \quad (147)$$

*Proof:* The proof is basically the same as that of Proposition 2 in [39]. Fix an arbitrary  $\delta > 0$ . Let  $\mathcal{H}_{n,\delta}^{\bar{A}\bar{B}} \subset (\mathcal{H}^{AB})^{\otimes n}$  and  $\mathcal{H}_{n,\delta}^{\bar{B}} \subset (\mathcal{H}^B)^{\otimes n}$  be the  $\delta$ -weakly typical subspace with respect to  $(\rho^{AB})^{\otimes n}$  and  $(\rho^B)^{\otimes n}$ , and let  $\Pi_{n,\delta}^{\bar{A}\bar{B}}$  and  $\Pi_{n,\delta}^{\bar{B}}$  be the projection onto those subspaces, respectively. There exists a  $\delta$ -independent constant  $c > 0$  such that we have

$$\begin{aligned} \text{Tr}[\Pi_{n,\delta}^{\bar{A}\bar{B}} (\rho^{AB})^{\otimes n}] & \geq 1 - \exp(-c\delta^2 n) \\ \text{Tr}[\Pi_{n,\delta}^{\bar{B}} (\rho^B)^{\otimes n}] & \geq 1 - \exp(-c\delta^2 n) \end{aligned}$$

for any  $\delta > 0$  and  $n$  [40]. Define

$$\begin{aligned} \tilde{D}_{n,\delta} &:= \text{Tr}[\Pi_{n,\delta}^{\bar{B}} \Pi_{n,\delta}^{\bar{A}\bar{B}} (\rho^{AB})^{\otimes n} \Pi_{n,\delta}^{\bar{A}\bar{B}} \Pi_{n,\delta}^{\bar{B}}] \\ \tilde{\rho}_{n,\delta}^{\bar{A}\bar{B}} &:= \frac{\Pi_{n,\delta}^{\bar{B}} \Pi_{n,\delta}^{\bar{A}\bar{B}} (\rho^{AB})^{\otimes n} \Pi_{n,\delta}^{\bar{A}\bar{B}} \Pi_{n,\delta}^{\bar{B}}}{\tilde{D}_{n,\delta}}. \end{aligned} \quad (148)$$

Due to Lemma 32, we have

$$\left\| (\rho^{AB})^{\otimes n} - \tilde{\rho}_{n,\delta}^{\bar{A}\bar{B}} \right\|_1 \leq 5 \exp\left(-\frac{c\delta^2 n}{4}\right) \quad (149)$$

in addition to

$$\tilde{D}_{n,\delta} \geq 1 - 2 \exp\left(-\frac{c\delta^2 n}{2}\right). \quad (150)$$

Let  $\Pi_{n,\delta}'^{\bar{B}}$  be the projection onto the subspace of  $\mathcal{H}_{n,\delta}^{\bar{B}}$  spanned by the eigenvectors of  $\tilde{\rho}_{n,\delta}^{\bar{B}}$ , corresponding to the eigenvalues not smaller than

$$\lambda^* := 2^{-n(S(\rho^B) + \delta)} \cdot \exp\left(-\frac{c\delta^2 n}{2}\right)$$

Define

$$\begin{aligned} D'_{n,\delta} &:= \text{Tr}[\Pi_{n,\delta}'^{\bar{B}} \tilde{\rho}_{n,\delta}^{\bar{A}\bar{B}}], \\ \hat{\rho}_{n,\delta}^{\bar{A}\bar{B}} &:= \frac{\Pi_{n,\delta}'^{\bar{B}} \tilde{\rho}_{n,\delta}^{\bar{A}\bar{B}} \Pi_{n,\delta}'^{\bar{B}}}{D'_{n,\delta}}. \end{aligned} \quad (151)$$

We then have

$$\begin{aligned} D'_{n,\delta} &= 1 - \text{Tr}[(I^{\bar{B}} - \Pi_{n,\delta}'^{\bar{B}}) \tilde{\rho}_{n,\delta}^{\bar{B}}] \\ &\geq 1 - \lambda^* \times \text{rank}[\tilde{\rho}_{n,\delta}^{\bar{B}}] \\ &\geq 1 - 2^{-nc\delta^2 \log e/2} \\ &\geq 1 - \exp\left(-\frac{c\delta^2 n}{2}\right), \end{aligned} \quad (152)$$

where we used the fact that

$$\text{rank}[\tilde{\rho}_{n,\delta}^{\bar{B}}] \leq \dim \mathcal{H}_{n,\delta}^{\bar{B}} \leq 2^{n(S(\rho^B) + \delta)}.$$

Therefore, due to the gentle measurement lemma, we have

$$\left\| \tilde{\rho}_{n,\delta}^{\bar{A}\bar{B}} - \hat{\rho}_{n,\delta}^{\bar{A}\bar{B}} \right\|_1 \leq 2 \exp\left(-\frac{c\delta^2 n}{4}\right). \quad (153)$$

From (149), (153) and the triangle inequality, we obtain

$$\left\| (\rho^{AB})^{\otimes n} - \hat{\rho}_{n,\delta}^{\bar{A}\bar{B}} \right\|_1 \leq 7 \exp\left(-\frac{c\delta^2 n}{4}\right). \quad (154)$$

From Definitions (148), (151) and Inequalities (150), (152), the maximum eigenvalue  $\lambda^+$  of  $\hat{\rho}_{n,\delta}^{\bar{A}\bar{B}}$  is bounded as

$$\begin{aligned} \lambda^+ &\leq \frac{2^{-n(S(\rho^{AB}) - \delta)}}{\tilde{D}_{n,\delta} D'_{n,\delta}} \\ &\leq \frac{2^{-n(S(\rho^{AB}) - \delta)}}{1 - 3 \exp\left(-\frac{c\delta^2 n}{2}\right)} \leq 2^{-n(S(\rho^{AB}) - 2\delta)} \end{aligned} \quad (155)$$

for sufficiently large  $n$ . By definition, we also have

$$\frac{1}{\lambda^+} \hat{\rho}_{n,\delta}^{\bar{A}\bar{B}}(V) \leq (\Pi^A)^{\otimes n} \otimes \Pi_{n,\delta}'^{\bar{B}}. \quad (156)$$

Let  $\{p(dV), V\}$  be an ensemble of unitaries on  $(\mathcal{H}^A)^{\otimes n}$  that satisfies (145), and define

$$\hat{\rho}_{n,\delta}^{\bar{A}\bar{B}}(V) := V^{\bar{A}} \hat{\rho}_{n,\delta}^{\bar{A}\bar{B}} V^{\dagger \bar{A}}. \quad (157)$$

As an ensemble average, we have

$$\bar{\rho}_{n,\delta}^{\bar{A}\bar{B}} := \mathbb{E}[\hat{\rho}_{n,\delta}^{\bar{A}\bar{B}}(V)] = (\pi^A)^{\otimes n} \otimes \hat{\rho}_{n,\delta}^{\bar{B}}. \quad (158)$$

Inequality (154) then implies

$$\begin{aligned} \left\| \bar{\rho}_{n,\delta}^{\bar{A}\bar{B}} - (\pi^A)^{\otimes n} \otimes (\rho^B)^{\otimes n} \right\|_1 &= \left\| \hat{\rho}_{n,\delta}^{\bar{B}} - (\rho^B)^{\otimes n} \right\|_1 \\ &\leq \left\| \hat{\rho}_{n,\delta}^{\bar{A}\bar{B}} - (\rho^{AB})^{\otimes n} \right\|_1 \leq 7 \exp\left(-\frac{c\delta^2 n}{4}\right), \end{aligned} \quad (159)$$

where the second line follows from the monotonicity of the trace distance. Due to (151) and (152), the minimum nonzero eigenvalue  $\lambda^-$  of (158) is bounded as

$$\lambda^- \geq \frac{\lambda^*}{d_A^n D'_{n,\delta}} \geq d_A^{-n} \cdot \lambda^*,$$

which leads to

$$\begin{aligned} \lambda &:= \frac{\lambda^-}{\lambda^+} \\ &\geq 2^{-n[\log d_A + S(\rho^B) - S(\rho^{AB}) + 3\delta]} \cdot \exp\left(-\frac{c\delta^2 n}{2}\right) \\ &= 2^{-n[I(A:B)_\rho + 3\delta]} \cdot \exp\left(-\frac{c\delta^2 n}{2}\right). \end{aligned}$$

Suppose  $V_1, \dots, V_N$  are unitaries that are randomly and independently chosen from an ensemble  $\{p(dV), V\}$ . Due to (156) and the operator Chernoff bound (Lemma 3 in [39]), we have

$$\begin{aligned} &\Pr\left\{\frac{1}{N} \sum_{i=1}^N \hat{\rho}_{n,\delta}^{\bar{A}\bar{B}}(V_i) \notin \left[(1-\epsilon_1)\bar{\rho}_{n,\delta}^{\bar{A}\bar{B}}, (1+\epsilon_1)\bar{\rho}_{n,\delta}^{\bar{A}\bar{B}}\right]\right\} \\ &= \Pr\left\{\frac{1}{N} \sum_{i=1}^N \frac{\hat{\rho}_{n,\delta}^{\bar{A}\bar{B}}(V_i)}{\lambda^+} \notin \left[(1-\epsilon_1)\frac{\bar{\rho}_{n,\delta}^{\bar{A}\bar{B}}}{\lambda^+}, (1+\epsilon_1)\frac{\bar{\rho}_{n,\delta}^{\bar{A}\bar{B}}}{\lambda^+}\right]\right\} \\ &\leq 2d_A^n d_B^n \exp\left(-\frac{N\lambda\epsilon_1^2}{2}\right) \end{aligned}$$

for any  $\epsilon_1 \in (0, 1]$ , which implies that

$$\begin{aligned} &\Pr\left\{\left\|\frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} \hat{\rho}_{n,\delta}^{\bar{A}\bar{B}}(V_i) - \bar{\rho}_{n,\delta}^{\bar{A}\bar{B}}\right\|_1 \leq 2\epsilon_1\right\} \\ &\geq 1 - 2d_A^n d_B^n \exp\left(-\frac{2^{nR}\lambda\epsilon_1^2}{2}\right) \end{aligned} \quad (160)$$

for an arbitrary  $R > 0$ . Substituting  $\epsilon_1 = 2^{-1-n\delta}$ , we obtain

$$\begin{aligned} &\Pr\left\{\left\|\frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} \hat{\rho}_{n,\delta}^{\bar{A}\bar{B}}(V_i) - \bar{\rho}_{n,\delta}^{\bar{A}\bar{B}}\right\|_1 \leq 2^{-n\delta}\right\} \\ &\geq 1 - 2d_A^n d_B^n \exp\left(-\frac{2^{n(R-2\delta)}\lambda}{8}\right). \end{aligned}$$

Therefore, if  $R$  satisfies

$$R > I(A:B)_\rho + 5\delta + \frac{1}{2}c\delta^2 \log e, \quad (161)$$

and if  $n$  is sufficiently large so that Inequality (155) holds and the R.H.S. in (160) is greater than 0, there exists a set of unitaries  $\{V_i\}_{i=1}^{2^{nR}}$  such that

$$\left\|\frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} \hat{\rho}_{n,\delta}^{\bar{A}\bar{B}}(V_i) - \bar{\rho}_{n,\delta}^{\bar{A}\bar{B}}\right\|_1 \leq 2^{-n\delta}. \quad (162)$$

Using unitaries in the set, construct a random unitary operation  $\mathcal{V}_n$  on  $A^n$  as (146).

The total error is evaluated as follows. From (154), (157) and the monotonicity of the trace distance, we have

$$\left\|\mathcal{V}_n((\rho^{AB})^{\otimes n}) - \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} \hat{\rho}_{n,\delta}^{\bar{A}\bar{B}}(V_i)\right\|_1 \leq 7 \exp\left(-\frac{c\delta^2 n}{4}\right). \quad (163)$$

Due to (159), (162), (163) and the triangle inequality, we obtain

$$\begin{aligned} &\left\|\mathcal{V}_n((\rho^{AB})^{\otimes n}) - (\pi^A)^{\otimes n} \otimes (\rho^B)^{\otimes n}\right\|_1 \\ &\leq 2^{-n\delta} + 14 \exp\left(-\frac{c\delta^2 n}{4}\right) \end{aligned}$$

for any  $R > I(A:B)_\rho$ ,  $\delta \in (0, 1]$  satisfying (161) and sufficiently large  $n$ . Thus we obtain (147). ■

## B. Proof of Theorem 26

We prove Theorem 26 by showing that  $M(U) \leq K(U)$  holds for any generalized Clifford operator  $U$ , which implies  $M(U) = K(U)$  due to Lemma 24. From Equality (43), we have

$$\begin{aligned} K(U) &= S(AR_A)_{\Psi_U} \\ &= S(AR_A)_{\Psi_U} + \log d - \log d \\ &= S(AR_A)_{\Psi_U} + S(R_B)_{\Psi_U} - S(B)_{\Psi_U} \\ &= S(AR_A)_{\Psi_U} + S(R_B)_{\Psi_U} - S(AR_A R_B)_{\Psi_U} \\ &= I(AR_A : R_B)_{\Psi_U}. \end{aligned}$$

Thus it suffices to prove that any  $R$  satisfying  $R > I(AR_A : R_B)_{\Psi_U}$  also satisfies  $R \geq M(U)$ .

Fix an arbitrary  $R > I(AR_A : R_B)_{\Psi_U}$  and choose sufficiently small  $\delta$  and sufficiently large  $n$ . Define

$$\begin{aligned} \vec{p} &:= (p_1, \dots, p_n) \in \{1, \dots, d\}^n \\ \vec{q} &:= (q_1, \dots, q_n) \in \{1, \dots, d\}^n \\ \sigma_{\vec{p}\vec{q}} &:= \sigma_{p_1 q_1} \otimes \dots \otimes \sigma_{p_n q_n}, \end{aligned}$$

and consider the ensemble of unitaries

$$\{1/d^{2n}, \sigma_{\vec{p}\vec{q}}\}_{\vec{p}\vec{q} \in \{1, \dots, d\}^{2n}}$$

on  $R_B^{\otimes n} = \bar{R}_B$ . Because of Schur's lemma, the ensemble satisfies

$$\frac{1}{d^{2n}} \sum_{\vec{p}\vec{q}} \sigma_{\vec{p}\vec{q}}^{\bar{R}_B} |\phi\rangle\langle\phi|^{\bar{R}_B} \sigma_{\vec{p}\vec{q}}^{\dagger \bar{R}_B} = (\pi_d^{R_B})^{\otimes n}$$

for any  $|\phi\rangle \in \mathcal{H}^{\bar{R}_B}$ . Therefore, due to Lemma 36, there exists a subset  $\{\vec{p}_k \vec{q}_k\}_{k=1}^{2^{nR}} \subset \{1, \dots, d\}^{2n}$  such that

$$\left\| \frac{1}{2^{nR}} \sum_{k=1}^{2^{nR}} \sigma_{\vec{p}_k \vec{q}_k}^{\bar{R}_B} (\Psi_U^{\otimes n})^{\bar{A} \bar{R}_A \bar{R}_B} \sigma_{\vec{p}_k \vec{q}_k}^{\dagger \bar{R}_B} - (\Psi_U^{\otimes n})^{\bar{A} \bar{R}_A} \otimes (\pi_d^{R_B})^{\otimes n} \right\|_1 \leq \epsilon_n, \quad (164)$$

where

$$\epsilon_n := 2^{-n\delta} + 16 \exp\left(-\frac{c\delta^2 n}{4}\right). \quad (165)$$

Without loss of generality, we assume that the basis  $\{|t\rangle\}_{t=1}^d$ , by which the generalized Pauli operators are defined as (48), is the Schmidt basis of  $|\Phi_d\rangle^{A R_A}$  and  $|\Phi_d\rangle^{B R_B}$ . That is, we assume that

$$|\Phi_d\rangle^{A R_A} = \frac{1}{d} \sum_{t=1}^d |t\rangle^A |t\rangle^{R_A}, \quad |\Phi_d\rangle^{B R_B} = \frac{1}{d} \sum_{t=1}^d |t\rangle^B |t\rangle^{R_B}.$$

A simple calculation then leads to

$$\begin{aligned} \sigma_{pq}^{R_B} |\Phi_d\rangle^{B R_B} &= (\sigma_{pq}^T)^B |\Phi_d\rangle^{B R_B} \\ &= \exp(2\pi i q p / d) \cdot \sigma_{-p, q}^B |\Phi_d\rangle^{B R_B}, \end{aligned}$$

where the superscript  $T$  denotes the transposition with respect to the basis  $\{|t\rangle\}_{t=1}^d$  defined by

$$X^T = \sum_{t, t'=1}^d \langle t | X | t' \rangle \cdot |t'\rangle \langle t|$$

for  $X \in \mathcal{L}(\mathcal{H}^B)$ . Therefore, for some phase  $\theta'_{pq} \in \mathbb{R}$  we have

$$\begin{aligned} \sigma_{pq}^{R_B} |\Psi_U\rangle^{A B R_A R_B} &= (U^{AB} \otimes \sigma_{pq}^{R_B}) |\Phi_d\rangle^{A R_A} |\Phi_d\rangle^{B R_B} \\ &= \exp(2\pi i q p / d) \cdot U^{AB} (I^A \otimes \sigma_{-p, q}^B) |\Phi_d\rangle^{A R_A} |\Phi_d\rangle^{B R_B} \\ &= e^{i\theta'_{pq}} (\sigma_{p'q'}^A \otimes \sigma_{r's'}^B) U^{AB} |\Phi_d\rangle^{A R_A} |\Phi_d\rangle^{B R_B} \\ &= e^{i\theta'_{pq}} (\sigma_{p'q'}^A \otimes \sigma_{r's'}^B) |\Psi_U\rangle^{A B R_A R_B}. \end{aligned}$$

Tracing out  $B$ , we obtain

$$\sigma_{pq}^{R_B} \Psi_U^{A R_A R_B} \sigma_{pq}^{\dagger R_B} = \sigma_{p'q'}^A \Psi_U^{A R_A R_B} \sigma_{p'q'}^{\dagger A},$$

which implies

$$\sigma_{\vec{p}_k \vec{q}_k}^{\bar{R}_B} (\Psi_U^{\otimes n})^{\bar{A} \bar{R}_A \bar{R}_B} \sigma_{\vec{p}_k \vec{q}_k}^{\dagger \bar{R}_B} = \sigma_{\vec{p}'_k \vec{q}'_k}^{\bar{A}} (\Psi_U^{\otimes n})^{\bar{A} \bar{R}_A \bar{R}_B} \sigma_{\vec{p}'_k \vec{q}'_k}^{\dagger \bar{A}}$$

for a subset  $\{\vec{p}_k \vec{q}_k\}_{k=1}^{2^{nR}} \subset \{1, \dots, d\}^{2n}$ . Using the subset, construct a random unitary operation  $\mathcal{V}_n$  on  $\mathcal{S}(\mathcal{H}^{\bar{A}})$  as

$$\mathcal{V}_n(\cdot) = \frac{1}{2^{nR}} \sum_{k=1}^{2^{nR}} \sigma_{\vec{p}'_k \vec{q}'_k}^{\bar{A}} (\cdot) \sigma_{\vec{p}'_k \vec{q}'_k}^{\dagger \bar{A}}.$$

We then have

$$\left\| \mathcal{V}_n((\Psi_U^{\otimes n})^{\bar{A} \bar{R}_A \bar{R}_B}) - (\Psi_U^{\otimes n})^{\bar{A} \bar{R}_A} \otimes (\Psi_U^{\otimes n})^{\bar{R}_B} \right\|_1 \leq \epsilon_n$$

from (164). Note that  $\Psi_U^{R_B} = \pi_d^{R_B}$ . Due to Lemma 19 and the proof thereof (see Appendix C-B), it follows that the state

$\mathcal{V}_n((\Psi_U^{\otimes n})^{\bar{A} \bar{R}_A \bar{R}_B})$  is  $\epsilon_n$ -recoverable from  $A R_A$ . Therefore, from Definition 9 and Theorem 11 in [14], we obtain  $R \geq M(U)$ . Note that the error vanishes exponentially to  $n$  due to (165). ■

## APPENDIX H

### PROOF OF LEMMA 29 AND 30

#### A. Proof of Lemma 29

The proof is based on an idea which is used in [41] to prove that information causality is satisfied in quantum mechanics.

For Inequality (49), define  $L^0 = \hat{K}^0 = \emptyset$ , and denote system  $B$  before the first step by  $B_0$ . By the data processing inequality and the chain rule, we have

$$\begin{aligned} &I(\vec{X} : B_\gamma, L^\gamma, \hat{K}^\gamma) \\ &\leq I(\vec{X} : B_{\gamma-1}, L^{\gamma-1}, \hat{K}^\gamma) \\ &= I(\vec{X} : B_{\gamma-1}, L^{\gamma-1}, \hat{K}^{\gamma-1}) \\ &\quad + I(\vec{X} : \hat{K}_\gamma | B_{\gamma-1}, L^{\gamma-1}, \hat{K}^{\gamma-1}) \\ &= I(\vec{X} : B_{\gamma-1}, L^{\gamma-1}, \hat{K}^{\gamma-1}) \\ &\quad + I(\vec{X}, B_{\gamma-1}, L^{\gamma-1}, \hat{K}^{\gamma-1} : \hat{K}_\gamma) \\ &\quad - I(B_{\gamma-1}, L^{\gamma-1}, \hat{K}^{\gamma-1} : \hat{K}_\gamma) \\ &\leq I(\vec{X} : B_{\gamma-1}, L^{\gamma-1}, \hat{K}^{\gamma-1}) + H(\hat{K}_\gamma) \\ &\leq I(\vec{X} : B_{\gamma-1}, L^{\gamma-1}, \hat{K}^{\gamma-1}) + \log |\hat{\mathcal{K}}_\gamma| \end{aligned}$$

for  $\gamma = 1, \dots, \Gamma$ . Hence we obtain

$$\begin{aligned} &I(\vec{X} : B_\Gamma, L^\Gamma, \hat{K}^\Gamma) \\ &= \sum_{\gamma=1}^{\Gamma} \left[ I(\vec{X} : B_\gamma, L^\gamma, \hat{K}^\gamma) - I(\vec{X} : B_{\gamma-1}, L^{\gamma-1}, \hat{K}^{\gamma-1}) \right] \\ &\leq \sum_{\gamma=1}^{\Gamma} \log |\hat{\mathcal{K}}_\gamma| = C_{\text{tot}}. \end{aligned}$$

For Inequality (50), observe that

$$\begin{aligned} H(\vec{X}) - H(\vec{X} | \vec{X}') &= I(\vec{X} : \vec{X}') \leq I(\vec{X} : B_\Gamma) \\ &\leq I(\vec{X} : B_\Gamma, L^\Gamma, \hat{K}^\Gamma) \leq C_{\text{tot}} \end{aligned} \quad (166)$$

due to the data processing inequality. By definition, we have

$$H(\vec{X}) = nR. \quad (167)$$

Fano's inequality [42] implies

$$H(\vec{X} | \vec{X}') \leq h(P_e) + nR P_e, \quad (168)$$

Substituting (167) and (168) to (166), we obtain (50). ■

#### B. Proof of Lemma 30

The proof of the first statement is based on a protocol proposed in [43]. Let  $B_1$  and  $B_2$  be  $d$ -dimensional quantum systems, and let  $\{\sigma_i^A\}_{i=1}^{d^2}$  be the set of generalized Pauli operators on  $\mathcal{H}^A$ . Define states

$$|\Phi_i\rangle^{A B_1} := \sigma_i^A |\Phi_d\rangle^{A B_1} = (\sigma_i^T)^{B_1} |\Phi_d\rangle^{A B_1} \quad (169)$$

and

$$\begin{aligned} |\Psi_{U,i}\rangle^{AB_1BB_2} &:= U^{AB} |\Phi_i\rangle^{AB_1} |\Phi_d\rangle^{BB_2} \\ &= (\sigma_i^T)^{B_1} |\Psi_U\rangle^{AB_1BB_2} \end{aligned} \quad (170)$$

for  $i = 1, \dots, d^2$ . Let us introduce notations

$$\begin{aligned} \vec{i} &:= i_1 \dots i_n \in \{1, \dots, d^2\}^n, \\ \sigma_{\vec{i}} &:= \sigma_{i_1} \otimes \dots \otimes \sigma_{i_n} \\ |\Phi_{\vec{i}}\rangle^{\bar{A}\bar{B}_1} &:= |\Phi_{i_1}\rangle^{AB_1} \otimes \dots \otimes |\Phi_{i_n}\rangle^{AB_1}, \\ |\Psi_{U,\vec{i}}\rangle^{\bar{A}\bar{B}_1\bar{B}\bar{B}_2} &:= |\Psi_{U,i_1}\rangle^{AB_1BB_2} \otimes \dots \otimes |\Psi_{U,i_n}\rangle^{AB_1BB_2}, \end{aligned}$$

and

$$\rho(\mathcal{U}_n, \vec{i}) := \mathcal{U}_n(|\Phi_{\vec{i}}\rangle^{\bar{A}\bar{B}_1} |\Phi_d^{\otimes n}\rangle^{\bar{B}\bar{B}_2}).$$

From (169) and (170), we have

$$\rho(\mathcal{U}_n, \vec{i}) = (\sigma_{\vec{i}}^T)^{\bar{B}_1} \rho(\mathcal{U}_n) (\sigma_{\vec{i}}^*)^{\bar{B}_1}$$

and

$$|\Psi_{U,\vec{i}}\rangle^{\bar{A}\bar{B}_1\bar{B}\bar{B}_2} = (\sigma_{\vec{i}}^T)^{\bar{B}_1} |\Psi_U\rangle^{\bar{A}\bar{B}_1\bar{B}\bar{B}_2},$$

where we defined

$$\sigma_{\vec{i}}^* := (\sigma_{\vec{i}}^T)^\dagger.$$

Therefore, due to the unitary invariance of the fidelity (see Equality (69)), Condition (51) implies

$$F(\rho(\mathcal{U}_n, \vec{i}), |\Psi_{U,\vec{i}}\rangle) \geq 1 - \epsilon,$$

which leads to

$$F(\rho(\mathcal{U}_n, \vec{i})^{\bar{B}_1\bar{B}\bar{B}_2}, \Psi_{U,\vec{i}}^{\bar{B}_1\bar{B}\bar{B}_2}) \geq 1 - \epsilon \quad (171)$$

for any  $\vec{i}$  by taking the partial trace.

Due to Schur's lemma, we have

$$\frac{1}{d^2} \sum_{i=1}^{d^2} \sigma_i^A |\Phi_d\rangle \langle \Phi_d|^{AB_1} \sigma_i^{\dagger A} = \pi_d^A \otimes \pi_d^{B_1}.$$

Thus the average state of  $|\Psi_{U,i}\rangle$  with respect to the uniform distribution  $p_i = 1/d^2$  ( $i = 1, \dots, d^2$ ) is given by

$$\begin{aligned} \bar{\Psi}_U^{AB_1BB_2} &:= \frac{1}{d^2} \sum_{i=1}^{d^2} \Psi_{U,i}^{AB_1BB_2} \\ &= \pi_d^{B_1} \otimes U^{AB} (\pi_d^A \otimes \Phi_d^{BB_2}) U^{\dagger AB}. \end{aligned}$$

The reduced state of  $\bar{\Psi}_U$  on  $B_1BB_2$  is

$$\begin{aligned} \bar{\Psi}_U^{B_1BB_2} &= \pi_d^{B_1} \otimes \text{Tr}_A[U^{AB} (\pi_d^A \otimes \Phi_d^{BB_2}) U^{\dagger AB}] \\ &= \pi_d^{B_1} \otimes \text{Tr}_{AB_1}[U^{AB} (\Phi_d^{AB_1} \otimes \Phi_d^{BB_2}) U^{\dagger AB}] \\ &= \pi_d^{B_1} \otimes \sum_{s=0}^{d^2-1} c_s^2 F_s^B |\Phi_d\rangle \langle \Phi_d|^{BB_2} F_s^{\dagger B}, \end{aligned}$$

where  $c_s$  and  $F_s$  are defined by (8). The von Neumann entropy of states  $\Psi_{U,i}^{B_1BB_2}$  and  $\bar{\Psi}_U^{B_1BB_2}$  are then

$$S(B_1BB_2)_{\Psi_{U,i}} = S(A)_{\Psi_{U,i}} = \log d$$

and

$$S(B_1BB_2)_{\bar{\Psi}_U} = \log d + K(U),$$

respectively, where the latter follows from (10) and the orthonormality of  $\{F_s|\Phi_d\rangle^{BB_2}\}_s$ . Thus the Holevo information ([44], [45]), corresponding to the signal states  $\{\Psi_i^{B_1BB_2}\}_{i=1}^{d^2}$  and the uniform distribution  $p_i = 1/d^2$  ( $i = 1, \dots, d^2$ ), is given by

$$\begin{aligned} \chi(\{p_i, \Psi_i^{B_1BB_2}\}) &= S(B_1BB_2)_{\bar{\Psi}_U} - \frac{1}{d^2} \sum_{i=1}^{d^2} S(B_1BB_2)_{\Psi_{U,i}} \\ &= K(U). \end{aligned}$$

Due to the Holevo-Schumacher-Westmoreland theorem [44], [45], for any  $\epsilon > 0$  and sufficiently large  $n$ , there exists a subset  $\mathcal{C}_n \subset \{1, \dots, d^2\}^n$  of cardinality  $n(K(U) - \epsilon)$ , such that all elements in the set

$$\{\Psi_{U,\vec{i}}^{\bar{B}_1\bar{B}\bar{B}_2}\}_{\vec{i} \in \mathcal{C}_n}$$

are distinguishable up to a small error  $\epsilon$ . That is, there exists a measurement on  $\bar{B}_1\bar{B}\bar{B}_2$ , described by a set of measurement operators  $\{D_{\vec{i}}\}_{\vec{i} \in \mathcal{C}_n}$ , such that we have

$$P_e := \frac{1}{|\mathcal{C}_n|} \sum_{\vec{i} \in \mathcal{C}_n} \left(1 - \text{Tr} \left[ D_{\vec{i}} \Psi_{U,\vec{i}}^{\bar{B}_1\bar{B}\bar{B}_2} \right] \right) \leq \epsilon$$

Consider the following protocol in which Alice transmits  $n(K(U) - \epsilon)$  bits of classical message to Bob by  $\mathcal{U}_n$ , assisted by shared entanglement:

- 1) Alice and Bob initially share  $|\Phi_d\rangle^{AB_1} |\Phi_d\rangle^{BB_2}$ , where  $B_1$  and  $B_2$  are additional quantum registers that Bob has.
- 2) To send a message  $k \in \{1, \dots, 2^{nR}\}$  where  $R = K(U) - \epsilon$ , Alice chooses  $k$ -th element  $\vec{i}^k = i_1^k \dots i_n^k$  in  $\mathcal{C}_n$ , and applies  $\sigma_{\vec{i}^k}$  on  $A^n$ .
- 3) Alice and Bob apply  $\mathcal{U}_n$ .
- 4) Bob performs a measurement on  $\bar{B}_1\bar{B}\bar{B}_2$  described by  $\{D_{\vec{i}}\}_{\vec{i} \in \mathcal{C}_n}$ .

The state after Step 3) is equal to  $\rho(\mathcal{U}_n, \vec{i}^k)^{\bar{A}\bar{B}_1\bar{B}\bar{B}_2}$  for each  $k$ , which satisfies

$$\left\| \Psi_{\mathcal{U}_n, \vec{i}^k}^{\bar{B}_1\bar{B}\bar{B}_2} - \Psi_{U, \vec{i}^k}^{\bar{B}_1\bar{B}\bar{B}_2} \right\|_1 \leq 2\sqrt{\epsilon}$$

due to (171) and (70). Therefore, due to (55), the average error in transmitting the message is bounded above as

$$\begin{aligned} P'_e &:= \frac{1}{|\mathcal{C}_n|} \sum_{\vec{i} \in \mathcal{C}_n} \left(1 - \text{Tr} \left[ D_{\vec{i}} \Psi_{\mathcal{U}_n, \vec{i}}^{\bar{B}_1\bar{B}\bar{B}_2} \right] \right) \\ &= \frac{1}{|\mathcal{C}_n|} \sum_{\vec{i} \in \mathcal{C}_n} \left(1 - \text{Tr} \left[ D_{\vec{i}} \Psi_{U, \vec{i}}^{\bar{B}_1\bar{B}\bar{B}_2} \right] \right) \\ &\quad + \frac{1}{|\mathcal{C}_n|} \sum_{\vec{i} \in \mathcal{C}_n} \text{Tr} \left[ D_{\vec{i}} (\Psi_{U, \vec{i}}^{\bar{B}_1\bar{B}\bar{B}_2} - \Psi_{\mathcal{U}_n, \vec{i}}^{\bar{B}_1\bar{B}\bar{B}_2}) \right] \\ &\leq \epsilon + 4\sqrt{\epsilon} \leq 5\sqrt{\epsilon}, \end{aligned}$$

which completes the proof. ■